YANJIU FANGFA YU FANGFALUN CONGSHU

GONGGONG ZHENGCE
NEIRONG FENXI
FANGFA
LILUN YU YINGYONG

万卷方法/研究方法与方法论丛书
WANJUAN FANGFA
YANJIU FANGFA YU FANGFALUN CONGSHU

公共政策内容分析方法：理论与应用

■李 钢 蓝 石 等编著



Y A N J I U F A N G F A Y U F A N G F A L U N C O N G S H U

□封面设计：黄河封面工作室

GONGGONG ZHENGCE
NEIRONG FENXI
FANGFA
LILUN YU YINGYONG

这是一本介绍公共政策内容分析方法的作品，全书选取了20世纪40年代至今，国际上在不同时期用内容分析方法于政策分析的著名案例。内容涉及国际政治危机中的政策策略、美国总统竞选过程中新闻媒体对于竞选政策的报道特征、内容分析在教育政策分析中的运用、对于中国教育政策执行过程群众自发诉求反映的定量分析,以及对于美国学校办学理念的语义规律分析。

本书通过个案分析方法力图反映出内容分析方法在公共政策分析中成功应用的分析技巧。

本书是公共管理专业、教育与社会政策分析专业、新闻传播专业的教师和研究生进行政策分析方法学习中比较有用的参考书。

ISBN 978-7-5624-3850-2



定价：25.00元

Y A N J I U F A N G F A Y U F A N G F A L U N C O N G S H U

GONGGONG ZHENGCE
NEIRONG FENXI
FANGFA
LILUN YU YINGYONG

万卷方法/研究方法与方法论丛书

WAN JUAN FANGFA

公共政策内容分析方法：理论与应用

■编　著　李　钢　蓝　石　江雪梅
夏广志　徐志勇　邓春芳

重庆大学出版社

图书在版编目(CIP)数据

公共政策内容分析方法:理论与应用/李钢,蓝石等编著.—重庆:重庆大学出版社,2007.1

(万卷方法.研究方法与方法论丛书)

ISBN 978-7-5624-3850-2

Ⅰ.公... Ⅱ.①李...②蓝... Ⅲ.公共政策—分析方法 Ⅳ.D0

中国版本图书馆 CIP 数据核字(2006)第 125779 号

公共政策内容分析方法:理论与应用

李 钢 蓝 石 等编著

责任编辑:雷少波 版式设计:雷少波

责任校对:方 正 责任印制:张 策

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址:http://www.cqup.com.cn

邮箱:fxk@cqup.com.cn(市场营销部)

全国新华书店经销

重庆升光电力印务有限公司印刷

*

开本:787×1092 1/16 印张:10.5 字数:167千 插页:16 开 2 页

2007 年 1 月第 1 版 2007 年 1 月第 1 次印刷

印数:1—3 000

ISBN 978-7-5624-3850-2 定价:25.00 元

作者简介

李钢 山西人，教授，管理学博士，现执教于北京信息科技大学。获北京师范大学教育经济与管理专业博士学位，曾在清华大学21世纪发展研究院/公共管理学院专职做公共政策与管理方向博士后研究，并游学于美国芝加哥大学（UC）、北伊利诺大学（NIU）、西北大学等，进行公共教育政策执行绩效评价和高等工程教育管理方面的访问研究。著有《话语、文本与中国国家教育政策分析》、《高等院校发展战略转型》等教育政策研究专著。主译《创建有效的学习型学校》、《社会科学思维方法论基础》[①]等作品。现已主持承担完成中国博士后科学研究基金项目一项，参与完成国家级、省部级科研项目多项，并在《教育研究》、《中国高教研究》、《教育理论与实践》、《行为科学》、《中国大学教学》、《自然辩证法研究》、《生产力研究》等刊物上发表研究论文20余篇。目前主持承担中国国家自然基金委管理科学学部面上基金项目一项。李钢博士目前的研究方向主要是：公共政策分析与管理研究方法论、中国国家教育政策分析、高等教育发展战略、管理思想史。

蓝石 上海人，获得美国北伊利诺州UIUC大学领导科学与组织行为学专业方向博士学位。现任美国芝加哥地区 DeVry 大学计算机与电子工程学院院长，是美国 ABET（美国工程技术教育认证委员会）芝加哥地区的成员，IEEE 组织成员。蓝石博士目前的研究领域主要是：中美工程教育人才培养模式比较、中美工程教育管理跨文化比较研究。

---

① 该书中文译本作为"万卷方法"的一种，将由重庆大学出版社于2007年3月出版。

前　言

公共政策是人类社会发展到一定阶段后政治体系制度化运作的产物,也是政府调控和管理社会的规范和行动准则。公共政策分析方法是认识公共政策的重要工具。从分析方法的行为来看,有关公共政策的分析方法可以被分为两大范畴:定性的分析与定量的分析,这两种政策分析方法的范式互相交织使用,贯穿于社会公众对于公共政策认识理解过程当中。本书着力介绍对公共政策的内容进行定量分析的方法。这种对政策定量分析的方法是从有关政策主题信息的编码分析入手,进行政策信息语义的相关分析。这种方法也被称为政策的“内容分析法”或者政策“话语分析法”。

笔者于公共政策研究,尤其是教育政策分析过程中感受到,要想提高公共政策研究的水平,加强公共政策分析的研究方法积累至关重要。建构科学的政策研究方法论体系,无疑是在建构政策科学发展的基础。所以我花了两年多时间,并在游学美国的蓝石博士友情相助下,精心收集了美国学者若干篇用内容分析方法进行政策分析的文章,再加上我们自己的两篇用内容分析方法所做的研究论文,结集出版以飨读者。作为一本专题讨论政策研究方法的作品,本书的出版或许能够为公共政策的研究者提供一些可资参考的案例。

中国公共政策分析急需方法的提升

目前我国公共政策研究正面临着学科建设和研究范式的转型,主要表现在:随着公共政策研究水平的不断提高,公共政策分析人员已经不再满足于制造一系列“宏大的话语概念”来鼓吹一些“世

界级”的“时髦”的“热点论题”。因为，公共政策发展实践不断要求有关政策的研究者要拿出有说服力的实证分析方法，对于政策发展的实际过程提供更为细致、精确和定量化的分析，以利于公共政策的决策者和制定者能够正确地把握施政方针运行的未来趋势。

另一方面，公共政策分析领域中研究方法的贫困，又正在影响到社会各界对于公共政策分析整体学科的公信力。公共政策分析方法论研究成果缺乏，导致社会公众对公共政策分析方法的价值不够理解、产生疑问。一些来自公共管理领域内部的、狭隘的学科认识和偏见仍在轻视研究方法，各种研究基金“申请指南”中，原本是推动政策科学向高水平发展“质量保证手段”的研究方法，却往往都被安置于“研究指南”中次要的边缘角落！而实践中，有关公共政策分析方法的研究更是比不上炒作一个“新概念”那样容易获得人们的青睐。

政策分析方法的理论具有长效性与通识性。因为，公共政策的主题可以与时俱进地变化，政策分析方法理论却是永恒长存的。只要有公共政策存在，就会有相关的政策分析方法相助相伴，没有科学的政策分析方法支持，就不可能有真正学理意义上的政策分析研究存在。哪里有政策研究，哪里就有政策分析方法存在。政策分析方法只会随着政策的消亡而失去其研究的价值。政策分析标准和政策分析范式经过长时间的考验，才能体现出方法论的科学价值。

我们通过对公共管理比较先进的国家公共政策理论及实践的学习、体验和研究，并通过考察国内外公共政策发展过程发现：公共管理和公共政策研究水平的高与低，就必然会相应地表现为公共政策分析方法理论的丰富与贫困；对于政策分析方法多样性的重视与探索，就一定会促进政策决策研究水平的提高，从而最大程度地影响到政策决策的科学性。政策分析方法在推进政策研究领域和全社会理解和认识公共政策中起着至关重要的作用。这是发达国家从长期社会政策制定实践的经验和公共管理改革的教训中摸索出来的道理。所以，我们需要汲取发达国家在公共政策研究方面的各种方法技术，来服务于我国现代社会的公共政策学科建设。

基于上述考虑，我于策划编撰本书时，就力戒概念空谈的陈俗，以案例分析为主，专注于分析公共政策的真实过程。为此全书共选取了中国、美国的政策分析学者运用内容分析方法专题研究公共政策的研究论文若干篇。这些论文包括了有关国际政治危机中的政策策略信息编码定量分析；美国总统竞选过程中新闻媒体对于竞选

政策报道特征的定量分析;内容分析在教育政策定量分析中的运用;新世纪初,中国教育政策执行过程中,群众自发的政策诉求情况反映的定量分析;以及对于大量调查中得到的众多学校办学理念的语义规律分析,揭示出办学理念与学校教育方针层次之间的内在关系等方面的研究论文。论文分析的政策主题是显得有些发散,但是政策分析方法理论却是异乎寻常地集中,即专题集中于如何对公共政策进行内容分析。我相信认真的读者可以从阅读本书后体会到,自20世纪后半期以来有关公共政策分析研究方法发展变化的风格,尤其是有关政策内容分析方法理论演进的大致图景。

何为公共政策内容分析方法

有关公共政策分析的方法论一直是体现政策的决策者、政策的认知者和政策的执行者决策、理解和阐释公共政策的认识途径和执行水平所在。政策的分析方式也由于政策本身的多样性,以及政策所涉及的社会利益群体领域的多样性,而被用多种学科的知识所解释和分析。本书所辑的政策分析方法论文就是专门基于"文本"的公共政策分析方法。这些论文中的研究方法所体现出的观点其共识之处在于,尽管公共政策过程要涉及决策者、执行者和其余众多的社会资源,但是由于政策行为的规范性,在政策过程的每一个阶段都必然会有相应的政策文件被保存下来。正是由于公共政策决策是社会不同利益群体围绕政治意识形态力量进行协商达成均衡的体现。记录政策的文本是分析政策信息的基本出发点和真实凭证。由此可知,政策文件以及与政策相关的文本是政府政策行为的反映,文本的语义则是记述政策意图和政策过程尤为有效的客观凭证。

内容分析方法(content analysis)是本文郑重宣传的政策分析方法主干。内容分析法专指对于文献内容进行系统的定量与定性相结合的一种语言分析方法。其目的是要分析清楚或者说要测度出文献中有关主题的本质性的事实及其关联的发展趋势。这种分析方法的基本单元是可公开看到的文本(text)。由此,内容分析方法被誉为可以"从公开中萃取秘密"(卢泰宏,1997)。内容分析法是一个概念发源比较古老、方法技术却不断在更新的分析方法体系。仅仅就有关文字的内容和符号层面的内容分析,就可以上溯到迄今5 000年前中国史前时期的"占卜"技术和古代先哲们在《周易》中有关"祭祀"符号的考据。

有关文本记录符号的意义阐释,西方文化界也有先例,早在19世纪,瑞士著名语言学家索绪尔(Saussure. F)就曾提出在一般性地阅读文献的基础上,采用一些简单定量的统计方法来对文献中较长的"语序谱系"进行比较深入的分析和解释。他的这一设想由于当时没有适宜的计算技术方法支持,在很长的一段时间(大约有半个多世纪)没有实现方法技术上的突破。

随着数学分析理论和数理分析技术方法的不断进步,到20世纪上半叶,一些兼备有数学和人文知识的社会语言学家和心理学研究者在从事有关民族志研究和民族语言分布比较研究中逐步试验、实现和发展了索绪尔的这些设想,并且形成了有关社会语言信息的科学定量分析方法。第二次世界大战中,英国和美国的政治和新闻研究人员积极利用研究方法专长为盟军的战时情报服务[比如,哈罗德·拉斯韦尔等人(H. D. Lasswell, et al)的有关研究],这些学者成功地将内容分析方法运用于战时军事情报分析和战时宣传分析中,取得了非常重要的成果;内容分析方法一时被称为"战时传播研究"的重要组成。以至于在20世纪50年代至70年代期间,从事内容分析研究的人员及其研究成果也大多都被有关军方和情报部门所掌控。这也是在20世纪相当长的一段时期,社会公众很难看到运用内容分析方法进行系统研究公开发表论文的主要原因。由于信息认知不对称,以至于目前仍然还有人在"怀疑"内容分析方法究竟能不能够作为政策分析的一种分析方法而独立存在。

随着世界政治格局从"两极对峙"转向"多边发展",尘封于"深宫密闱"中的档案逐渐大白于社会公众,有关公共政策分析研究的禁忌疆界随之逐渐解禁开通,对于文献的内容分析方法也随之被运用于更为广阔的领域。甚至有人宣称在21世纪这一方法将要随着社会哲学的中心向"语言哲学"迁移,内容分析方法的运用范围将要"几乎遍及社会科学的所有研究领域"。

自20世纪后半期到21世纪初,公共政策的研究方法伴随着公共政策的发展,以及公共政策知识体系的不断复杂丰富而变得多样化、复杂化和具体化。因此,本书在收集有关的研究论文中,力图反映出内容分析方法在公共政策分析中的主题针对特性和方法普适功能。为此,全书选取了自20世纪40年代至今不同时期内容分析方法在政策分析中的研究案例。通过对有关政策主题个案分析方法的介绍,努力表现出进入20世纪后期,内容分析方法在公共政策分析评价中的发展趋向。

关于本书的编写

本书编写过程中通篇努力贯彻了三个原则:其一,政策分析方法的可操作性与可宣传推广价值;其二,不同语料分析案例的比较研究和对比分析;其三,内容分析方法适用的"长时段"历史分析和现实题材广谱分析运用相结合,以及分析的政策主题的即时性和趣味性。

在编撰本书过程中,我们投入了大量时间精力,力所能及地将论文中涉及的难度较高的数学编码公式的推导过程,进行了通俗化的技术加工剪接。这样读者在阅读中从逻辑上几乎感觉不到数据的枯燥,相反地,会感觉到与编码数据相关的主题带有数字语义上引人入胜的趣味性,如若不信的话,你可以翻阅本书的任意一章读一读来体味对政策进行内容分析的乐趣!

为了帮助读者尽快地进入政策分析的相关语境,以增加作品阅读的通俗性,有利于掌握这一有意义的政策分析方法。我采纳了出版社编辑的建议,查阅了有关政策分析的资料,精心于每一部分的开头写了"导读"和"术语解释"两部分内容,作为读者阅读正文前的背景知识路标。在此,我衷心感谢重庆大学出版社"万卷方法"丛书的策划人雷少波先生于百忙之中非常专业地审校了全书的文稿,提出了非常重要的修改建议。

我相信读者于阅读本书中将能够理论联系实际,批判地选取方法论的特点,开阔政策分析的视野。书中所涉及的有关公共政策制定的价值观分析、政策执行过程的量化评价,以及公共政策传播的主题意识分析等内容都是国际、国内研究界比较关注的有意义的信息。所选论文在内容分析方法原理的操作叙述细致得当,具有方法论运用方面的参考价值。同时,由于所集论文的多位作者政治信仰各异,政策观点不同,作者在此郑重提醒读者,阅读本书时,也应当善于运用内容分析方法,对其中一些不适宜于我们发展的消极观点予以认真地甄别和严肃地批判。

我相信这本书将对于那些真正将志趣投在公共管理专业、教育与社会政策分析专业、新闻传播专业的本科高年级学生、研究生和教师学习掌握科学的政策分析方法是有用的。

李　钢

2006 年 1 月于清华园

目 录

第 I 部分

政策内容分析的方法和原理

本部分中，我们将所涉及的这种面向文本的分析评估方法定义为“内容分析法(content analysis)”。这种方法是通过一系列的转换范式将非结构化文本中的自然信息转换成为可以用来定量分析的结构化的信息形态。因此，内容分析方法也被称为一种分析文本材料的结构化方法。本部分内容节选自美国国家总评估办公室(General Accountability Office，以下简称为 GAO)在培训政策分析专家时进行政策内容分析的教程。他们的训练计划要求未来的政策分析专家通过阅读和学习本文，应当能够对内容分析方法及其原理的基本概念和分析过程框架有一个比较清晰的了解。同时，掌握此法还有助于辨别内容分析方法的适用条件和适用范围，以及被分析文本内容的真实性。

尽管我们认为，运用技术手段将定量分析用于文本分析是规范化的可操作分析方法，但这并不意味着内容分析方法能够适用于对所有的文本进行分析。本篇所介绍的这种分析技巧是比较适用于 GAO 工作中其职能所要求的方法。在第 1 章中，将对内容分析方法进行界定，并对于内容分析方法的适用范围进行了叙述。在第 2 章中，对于内容分析法的优势和劣势进行了分析，并且描述了此法在项目评估中的一些应用实例。在第 3 章和第 4 章中，举出两个比较精彩的分析案例来叙述这种分析方法的特色，即，内容分析方法在 20 世纪已经过去的“冷战”中，持有相互对立的政治意识形态双方是如何在敌对状态的政策信息全然不能有效沟通的情形下，却能有效地利用内容分析方法来对于敌对一方的相关政策信息的真实性进行有效的分析判断！目前世界已经进入以和平和发展为主流

的多极世界政治文化格局,公共政策信息的网络化和多极化交流在一定意义上将会促进内容分析在更多领域的全面展开。

本部分编写中所涉及的研究报告是为研究人员自学分析方法所设计编辑的。其中有关的注释贯穿于整个文本过程中,这就为阅读者解释一些疑难的问题和比较复杂的分析题目提供了更多的信息和便利。

其中的一些报告是基于对有关内容分析方法的大量论文和研究文献进行检索和查阅的基础上形成的,在这过程中,报告作者还访谈了大量使用内容分析方法进行信息分析的研究人员,对内容分析法的长处和不足进行了充分的讨论。

我们将内容分析法用于选拔那些有志于从事项目评审评估的审计人员和学术研究者。应当明确的是,内容分析法也具有一定的适用性。本部分的前两章是在威廉姆・卡特(William Carter,1989)作为一个专业研究者与 GAO 合作时所写论文的基础上形成的。又由特瑞沙・帕丝卡整理形成正式文本,由项目评估与方法审查处(PEMD)正式发布。其中的一篇最初于 1982 年出版了较早的文本,并且同时发行了三种不同文字的文本。目前再版的这个文本仅仅就以前的文本略做改动和编辑,包括添加了一些新的参考注释。第 2 章和第 3 章所形成时的那些年代已经一去不复返了,其中一些言辞或许在今天阅读的时候会感到在语义上有悖于中国目前改革开放的主流社会意识话语,通过结合语言原义分析话语,来还原社会语境的真实面貌,这正是内容分析方法的历史价值所在。希望读者本着掌握方法之目的,能够了解书中分析的要领,而不必在只言片语上去做别有用心的猜忌或在枝节末梢上做毫无意义的追究。

应当强调的是,本部分所涉及的"内容分析法"基本上取材于由美国项目评估与方法审查处公开发表的一系列方法读物之一。他们编撰这些文本的目的是旨在为那些从事项目审计的人员提供一些比较清晰易懂的有关评估的基本理念和方法入门的读物。此外,本部分还用附件的形式来解释这种方法的一些背景和专项分析过程的应用分析案例。需要告诉读者的是,与本部分的主题有关的其余内容还包括了因素分析法、评估设计方案、结构化的访谈法、问卷设计使用、统计的实例分析、案例研究评估等。本书出于精练和主题明确的需要没有将其一一枚举列出,这并不意味着这些方法是次要的和可有可无的,只因为它们中任何一种方法都可以独立发展成为一种有关公共政策分析的方法论体系。所以,这些资料有待于

需要有志于继续进行内容分析的读者在今后的研究中自行去查找、去学习、去体会和实验。

术语解释

[**话语**(discourse)]　话语就是以若干词组和短句按照一定的语义联系和语法结构组合而成的以表示一个完整语义的集合体，这个集合体可以是一个句子，也可以是一个词组或者短句，还可以是一些长句的集合。话语作为语言分析单位，最初的出现是为了分析日常交际信息的需要，在2000年以前的《英语词典》中discourse的释义就局限于"对话、交际"。在梅洛—庞蒂、维特根斯坦、利奥塔、德里达、拉康和福柯等人有关后现代文化哲学的研究中，共同点在于研究日常会话和交流信息的关键是分析"话语"以及"话语权"。福柯认为"话语"不仅仅是分析语言文本信息的比较独立和关键的单元，同时更重要的是"话语"往往是由一系列陈述构成的，它经过一些学术霸权群体的认可和述说而获得受众的信任和统治权，权力与话语是密不可分的，话语一经产生就立即受到若干权力形式的控制、筛选、组织和再分配。而且在某些条件下，话语本身常常转化为权力。公共政策就是政府权力的施行过程，分析话语以及话语权是分析公共政策的根本和关键。

[**文本**(file)]　文本在本书中就是指由有关的话语组织的阐述公共政策的语篇。作为官方的政策文本就是公共政策的文件以及阐释政策的各种权威化包装的解释文本(比如:流行于市面上的各种有关公共政策的白皮书、绿皮书、蓝皮书，以及各类有关政策的研究著述等)。

[**编码**(Coding)]　编码是内容分析法的基本运作方式。没有量化编码的内容分析根本不能算是真正的内容分析。编码就是将欲分析的对象内容，按照分析框架所设定的分类指标进行内容材料的分类，并且要将分类的过程按照各种指标和标准的要求进行量化标记。编码是内容分析搜索政策内涵问题的唯一量化手段。由于分析框架的预设是内容分析的难点，因此，依照对于分析材料的认识，编码方式可以分为完全依照被分析材料的自然属性所进行的编码规程，称之为"自然编码"；以及按照分析人员根据相似研究所得出的经验预设，采取前设的编码原则，进行

结构化的标准编码设计,这样的编码被称之为"结构化编码"。本书中第一部分所采用的编码以"结构化编码"为主。第二部分所采用的编码以"自然编码"为主。这两种编码方式各有优点和不足,但二者可以结合使用。如何具体采用以哪一种编码方式为主,这要取决于政策分析人员的实战经验,掌握科学编码的秘诀在于,政策分析人员必须亲自动手对大样本的语料逐一进行仔细反复的编码,而不假手于他人。因为,内容分析的许多重要结论正是政策研究者亲自动手进行耐心、认真、踏实的编码过程中感悟获得的!

[**美国国家总评估办公室**(**The U. S. A General Accountability Office, GAO**)] 该组织据称是为美国国会和美国人民效力的官方组织机构,该组织受国会指派来研究联邦政府的财政支出并对于一些公共政策的全过程立项进行专门地评估调查;该组织是无党派属性的独立的调查研究评估机构。美国国家总评估办公室的主要职责在于,研究联邦政府是如何依法使用纳税人的钱,并通过公共政策和管理竭诚地服务于社会的。该组织的工作是通过报告、通知、咨议等方式对国会和政府提供政策咨询。国会及其具体的职能机构(例如环境保护组织、国家安全委员会、国民健康服务委员会等)可以根据美国国家总评估办公室的分析报告,对于政府的行为职能及其行使公共政策行为等方面进行多种途径的监督和干预。美国国家总评估办公室评估联邦政府的政策项目、对于联邦政府的财政活动进行查账、进行立法和司法建议的调研评估、提供相关的政策建议。由于美国国家总评估办公室卓有成效的工作,对于美国国家司法体系和联邦政府的合法运行起到了有效的监督作用,为国家积累和聚集了大量的财富。为此,这个组织机构被美国社会公众广泛誉评为效忠国家的"职业调查队伍"和国会的"神圣看门狗"。

[**美国住房与城市化发展部**(**Housing and Urban Development, HUD**)] 它是联邦政府行政机关的一个部,始建于1965年,是作为林登·B·约翰逊总统提出的"伟大社会计划"的一部分而设立的政府部门,以解决都市化社会出现的日益复杂的问题。该部享有内阁地位,由一名部长领导工作。该部所属的主要机构包括联邦住房管理局、联邦保险局以及国家水灾保险计划处。并且,随着美国社会进一步城市化发展,该部的机构和职能也在不断地拓宽和调整,以发挥联邦政府在保证国民住房和城市协

调发展方面的政府主导作用。

[**项目评估与方法审查处**(Program Evaluation and Methodology, PEMD)]　它是隶属于 GAO 的一个工作相对独立的部门,它的职能是专门立项对于特定的公共政策项目执行过程进行特别审计和评估,同时也对 GAO 的其他部门的调研和审计核查提供方法理论方面的咨询。这个部门的工作直接对国会的司法委员会领导负责。

第 1 章[①]

内容分析的过程和步骤

GAO 的人员在工作中经常会接触到大量有关政策文字符号的文本。例如，各类工作报告、委托协议、会议纪要和综述、预评估报告等，其中包括大量有关政策信息的可分析的有用的文本材料，并且这些文本材料形式多样，难以按照常规的方法和程序进行量化分析处理。

分析这些复杂材料的一种有效方法是，设法将这些散乱的文本通过结构化的转换变成为可以分析的文本，并且将文本当中主要的信息用表列的形式显示出来，然后再将这些主要的信息发生的频率统计出来。以上这些操作的内容部分是国会规定 GAO 的工作职责制度所必须要求实现的，另外的研究部分是属于内容分析方法所要求的范畴。

例如，在美国住房与城市化发展部的评估体系中，有关评估的办法其用意在于建议项目官员是否具有实现工作和方法双重性以及它们的双关影响。GAO 的审计官员通过收集各种财务预算信息、结合访谈调查，查阅评估报告等，他们通过从美国住房与城市化发展部方面收集来的 38 个评估报告，以及来自两个办公机构的项目报告，开始确定了 31 项主要信息指标来评估城市和农村的居住情况的变化。并且通过这些调查检索“两办”的文献目录来甄别文

---

① 本章及第 2 章的内容编译自 GAO 于 1989 年 3 月出版的《内容分析：一种结构化分析文本的方法》(*Content Analysis: A Methodology for Structuring and Analyzing Written Material*) 中的前三章。此处使用已获得该组织负责政策项目评估与方法审查处助理评估主管依连诺 · 车姆莱斯基(Eleanor Chemlimsky)博士的授权许可。本部分各章由李钢、邓春芳编译，李钢审校。为方便读者阅览，编者对内容结构有调整。

本档案材料中内容交错重叠的部分。这样的案例调查及其有关的协调，其目的是为了简化文本的条目，有利于内容分析方法的施展。

除了要对文本材料进行内容概括，同时就所述问题做相关频数统计外，GAO 项目官员经常还被要求进行更为复杂的政策分析，要面对大量的文本进行文献内容综述，通过内容分析得出结论，并推测得出一些发展趋势。这些推测要留待实际时间来检验，或许还需要经过在不同的形势，在不同的人群中进行检验。尽管这样复杂类型的文本信息需要用计算机分析，但在有关计算机分析的文档中并不多见。然而，采用内容分析就可以将比较复杂、散乱的文件，通过编码转化成为结构化的可以借助于计算机进行处理的标准化的信息进行文本分析。

内容分析是一种以标准化的格式收集、组织信息资源的过程。在这个过程中，允许分析人员在编码和分析文本的特点上存在一定的分析误差。可以形成简单的格式以易于反映信息的概况和陈述频数统计。分析趋势若是描述频数上细微的差异则需要更为复杂的格式。

接下来所阐述的内容分析程序表述当中，我们将集中讨论如何简化和界定分析对象并对文本材料进行分析编码的技术问题，以及如何分析、发展和开发标准化的目录编辑方式，采用固定化的信度和效度较高的、可靠性强的检查方式来分析和解释被分析的信息，并且有效地报告分析的结果。

本章对内容分析的过程进行详细的阐述。其中表 1.1 中的第 1,2,6 步提供了甄别内容分析方法是否可以适用、什么材料应该被分析，以及如何分析和解释结果的原则和方法。其中在第 3,4,5 步，包含了选择分析单元、拓展编码目录，以及语料编码等对于内容分析至关重要和特殊的部分。这些内容旨在确保内容分析和解释更为精确和真实。

表 1.1　内容分析的步骤

1. 决定是否采用内容分析
2. 确定哪些文本应当被列于内容分析(视野)当中
3. 选取分析单元
4. 拓展编码的目录
5. 对文本编码
6. 分析和解释结果

决定是否采用内容分析

在第1步,分析专家应当就是否采用内容分析来考虑大量的确定因素,这关系到要考虑项目的真实可行性,数据获取的可行性,以及究竟有哪些内容分析方法可以采用。

目标界定(对象的界定)

目标应当被首先描述成为一系列容易被用词句来表达的问题,这些问题容易让人解读和回答。这些问题的构造应当基于一些可获得的数据和客观项目的需要。用词语来精确地描述所收集到的数据、所设计的分析目标,以及报告中所具有的特征问题。总之,内容分析法能够被用作解释和回答"什么"的问题,它不能完全解决"为什么"的问题。或许此法是帮助分析人员描述和总结文本材料的内容,识别其中作者的写作特征,或者体察文本带给作者的影响。

正如我们前面提到的,文本的内容是可以通过对于其中的主题和主要观点的列表计量得以被精确地陈述和掌握。作者的态度和写作意图也随着对于这些陈述和总结归纳而被有效测度。例如,假设分析者要测度现今有关老年人的各类问题以及效果。有关这些问题的开发的访谈内容分析,也许能够被用来确认他们有关生活的期望值和对待毫无安全保障生活的态度。内容分析在信息接收效率的测度方面也颇有优势。例如,美国广播的传播效果可以通过分析前苏联的新闻报纸以及广播媒体对美国广播的转播情况得以测度。

什么样的材料可用来分析

内容分析可以用于分析任何声像材料和文本材料,只要这些材料是通过可靠、真实渠道获得的。因此,内容分析被频繁地用来分析文本材料。文本材料甚至包括了一些传媒的记录、电视栏目、电影,以及相关的图片等。内容分析也被用于分析议会的声明、立法过程、司法条文,以及其他公共文件、工作报告、案例研究,报道、调查研究结论、新闻出版物、报刊、书籍、专栏文章和信函等。但是,讲演和讨论一般不能用来进行内容分析,除非这些讲演和讨论被转化成为纸质的文本。

进行内容分析之前,项目研究人员应当充分估量文本材料的质

量特征，即，要回答所收集的文本的信息量是否充分回答了所要研究的问题？若在分析前就对文本进行断章取义式的删节，其结果将导致分析结论的片面性和不准确，形成对问题认识的谬误。只有对于文本材料进行大量、充分的实证分析，其内容分析的结论才能够真正的可信。

能用作对照分析的材料种类

内容分析被用于在文本之间进行大量的对照分析。例如，进行文本分析的研究人员应用内容分析法从某一稳定的信息源获取文本进行分析；如，对从联邦某机构获得的文本进行比较长时间的主要观点变化分析，并对照分析不同群体、不同形势下的观点的变化，以及在同一份文本当中，对两个或更多命题或观点进行比较分析也是常用的方法。与此相应，来自于两个或更多的文献源的不同的观点或相同类型的观点和陈述也可以被用来进行对比分析。

内容分析语料的选择

对于被分析的文本数量巨大、研究耗资过甚时，采取适当的抽样是很有必要的。因此，在操作的第2步，分析人员面对浩瀚无穷的文献，应当采取科学的抽样方法，以确保分析的结论中可以有效和正确地概括整个样本的全貌。

通常选取典型文本作为内容分析的样本，例如，在一个有关食品商品商标案例的内容分析项目中，假定按照适当的要求，做一个有关商标设计方案变化评估的研究。那么，预计可能要访谈的样本至少要有500位被访谈者。依照顺序初步分析访谈笔录内容，然后可以选取十分之一的记录进行内容分析。项目分析人员可能要设计一个系统的样本获取有关过程的方案，并且通过其他途径获取的文本样本也将被利用。

选取分析的信息单元

在内容分析方法中，研究人员设计的分析单元被称之为“编录单元(coding units)”。将上下文按照语境进行单元化的编录。在第3步中，内容分析被限定在这样的一类文本当中。这些文本的内容可以通过对字、词和句子进行编录和分类来获得。内容分析的单

元就是由这种字、词汇和句子编录的最小的信息单元。这些信息单元的语义有时并不能够保持一成不变,需要结合内容分析的需要进行适当地调整。

由于对那些长篇大论的文本难以迅速地进行精确的内容分析计量。因此,在选取信息单元时,篇章、段落、小节,以至于句子就能够作为较为合适的选择。分析单元的选取和细化,对于甄别内容上的细微变化非常有益。例如,有关会议的正式文本是否可以用来进行内容分析,取决于会议的参加者所发表的有关会议的各种类型的支持的或者是反对的观点,以及这些形形色色的会议论点所涉及的内容范围。在这个案例中,分析者可能选取句子作为分析的单元,因为整个会议的陈述观点太长或太五花八门。会议上什么样的观点都可以提出来,也包括相互之间带有冲突和对立的语义信息。譬如说,有一个讲演者在演说的过程中,可能一开始在抨击某一观点,然后到了讲演的最后却转过弯来赞成这一观点。识别这种具有戏剧性转换的话题迁移,分析人员就需要检验每一个小的分析单元,诸如句子。

编录单元在文本材料的分析框架中属于比较特别的编目部分,它可以是单词及其集合(诸如一些术语)、句子、段章,或者是整个的文本,但是这些都可以放在文本单元之内。在我们前面所提到的美国住房与城市化发展部的一些研究中,分析研究人员使用了一组词汇来表达对于有关论题讨论的分析编录单元。他们的语境化单元分析也被用于评估研究。

不断拓展的编码种类

编码种类显示出编录单元拓展的结构以及趋势。在第4步中,格式化的编码种类是确定内容分析的关键。白瑞瑟(Berelson,1952)作为一个较早涉足于内容分析领域的研究者,曾就编码的重要性慎重地强调:

> 内容分析的成败关键几乎全系于对其内容进行编码,专门的研究业已证实了编码范围需要提早确立,此举将十分有利于辨清和查找内容当中固有的语义问题。

表1.2列出编码可能遇到的标准化的规则,遵守这些规则有利于分析人员系统和客观地看待文本内容中,哪些不适用于统计计算,哪些对于结果的分析具有关键性的影响。

表 1.2　内容编码要求

1. 编码要保持尽量完美：即在一个编码系统中能够将所有相关的条目都包括在其中。
2. 目录之间要相互排斥：即不允许用两个以上编码代表同一个目录。
3. 编码之间保持独立：一个编码单元的设定不应当受到其他单元的影响。

编码格式化

编码可以用许多方式使之格式化。常用的编码格式化方法有分组法、刻度法、矩阵分析法等。

采用格式化的编码可以提高编码的有效性，尤其是在编码的种类形态庞大繁杂的时候更是如此。例如，在美国住房与城市化发展局提供的案例中，分析人员将观点进行分类编目，将31个观点分成三大类编目，并且具体化到将房屋的分布、街道的划分、公共房屋的现代化程度等都列于一个所谓的“房屋规划发展计划”的条目当中。采用了刻度法来划分欲分析信息的重要性等级。在美国住房与城市化发展局提供的案例中，由于分析人员已经对于要测度的问题和将要被测度的目标群体都有一个清醒的认识，因此，他们就采用了刻度法，将对问题的态度倾向划分为“支持”、“不参与”、“反对”。

矩阵分析法在分析人员寻找那种涉及过去和现在的多重变量的信息研究情况时，就可以将以上讨论过的分组法和刻度法组合形成一种矩阵态的分析方法（如表1.3所示）。

表 1.3　Ⅰ级目录标准化的编目矩阵

观点主题内容	观点的认同态度等级		
	支持	不参与	反对
Ⅰ级：房屋规划发展计划			
A　街道的划分			
B　房屋的分布			
C　公共房屋的现代化程度			

定量测度指标

编码常用于有关量化水平的测度，即有关距离、频数和密度等

的测量。为了说清量化分析与其相互关联的结构化编码分类的区别,我们将选取一个有关限制手枪使用的立法案例来综合分析。在这个案例分析中,研究人员实证分析的信息文本主要来自于报刊上的文章、公开的文件以及议会调查访谈后可以公开的报告文本。

研究人员在综合分析测度那些报刊文章所提出的支持或反对的论点的影响效果时,并不一定要严格恪守量化分析的所有法则。研究人员采用量化分析只是要比较准确地估计到那些对限制携带手枪持赞成或反对态度的观点发生次数的对比情况。对于样本中的每一份报纸中的每一个观点,研究人员都要考虑到所有有关的新闻报道中的专栏的篇幅长短,以及统计出对每一个立法背景进行分析的文章篇幅的大小。同时辅之以对每一份报纸的名称、属性、日期进行编码。分析人员可以通过农村和城市对于主题的不同态度随时间变化的情况来测度有关这项手枪限制立法事件发展的趋势(结果参见表 1.4)。

表 1.4　编码测量—主题态度关联表

标准化的编码测量空间			样本篇幅中对于主题态度的测量		
报刊新闻	时间	属地	支持	反对	不参与
Times《纽约时报》	1981-11-12	城市	4	0	2
Examiner《探索者报》	1981-11-18	农村	0	5	2

以上介绍的方法测量速度快且相对容易,但在分析上也仅仅提供了一个比较总体和笼统的印象。要进一步施展量化分析,研究人员需要针对性地预设条件,对于在大量的样本中发现的差异性进行细致分析,这样做是为将来分析主题的重要性和重点语义的指向提供充分的证据。

接下来进行的更为深入的量化分析中,研究人员将对编录单元出现的频次进行编码,根据记录每一个陈述或者观点在文本出现的次数来进行编码。

如表 1.5 所示,标准化测量频数的方法比较简单,但有时测量形式表现的非常复杂(如表 1.6)。繁简程度取决于项目所需求的信息测度的要求。

表 1.5　观测文本观点陈述测量的两类表格

（两个栏目中标准化的编码对于声明发生频数的测度）

格式 1

		表格中的项目栏目数量			
报刊	日期	属地	支持	反对	不参与
Times	1981-11-21	城市	2	0	1
Examiner	1981-11-21	农村	0	4	0

格式 2

报刊	日期	属地	言论主体	态度
Times	1981-11-12	城市	政党	支持
Times	1981-11-12	城市	编辑	支持
Times	1981-11-12	城市	国会参议员	不参与
Examiner	1981-11-18	农村	公民团体	反对
Examiner	1981-11-18	农村	国会政客	反对

表 1.6　有关特别提案的反映态度发生频数测量

测量态度分类			
有关控制手枪的提案	支持(01)	反对(02)	不参与/不表态(03)
严禁枪支销售	(01)		
禁止进口枪支零件	(02)		
手枪要登记	(03)		
严格控制购买手枪	(04)		
对使用枪支犯罪要严刑惩治	(05)		
对于已有的控制枪支制度要严厉执行	(06)		
其他观点	(07)		
编码形式		观点陈述	
信息源	时间	列	行
总统顾问团	1981-08-06	01	02
总统顾问团	1981-08-06	01	07
总统顾问团	1981-08-06	01	04

表 1.5 展示了有关控制枪支的不同态度和诉求观点发生频次

的两种测量列表。在以上两个表格中,请注意格式1类似于格式化的对一系列的文本在所给一段时间内出现的观点频数的简化测量表,但不能替代测量某些特定条文在特定时期出现的频次。格式2为分析人员提供了不同发言人在不同时间就所讨论主题发表观点的话语来源的不同属性、身份,以及精确罗列出了观点态度的差异性。

表1.6给出了对目录和分类编码设计中比较详尽的主题频率测量的格式,这种方法可被用来分析案例中——比如报纸、政府文本和面试记录稿这三类数据源中的信息。在表1.6中分析人员对控制手枪的具体提案意见做了分类描述。这些分类看法可被格式化地分成四组数字来描述,并且标出来提案内容分类的具体栏目的(行列)坐标位置。

为了说明以上列表操作分析的意义,可以引用1981年8月6日出版的《纽约时报》的陈述为例来验证这种分析,这些陈述可用表1.6显示。

> 总统顾问团中有8位成员建议立法:要求禁止枪支进口,要求公民报告盗窃或丢失枪支的情况,在确定购枪支者是否有犯罪记录以便确定其是否有权购买枪支前,留出一段观察和调查购枪者合法资格的时间。

立法禁止枪支进口的提案被编码放在第01行("支持"),第02列("禁止进口枪支零件"),因为第二条提案"要求公民举报盗窃或丢失枪支情况"没有列在具体提案建议清单中,所以被编码为"其他"(01,07)。

就测量的总体而言,当分析者实施测量频数的方法时,把两种假设合并在他们的研究设计中,首先,记录产生在文本叙述中的频数是一种有价值的或重要的有效数据。其次,分析者假设所有文本单位被赋予相等的权重,从而每一文本可以与其他另一文本进行直接比较。

在第三级水平上的量化分析中,分析员集中编码、统计频数。但相对集中的测量权重系数调整每一个被编码的陈述或观点,这种测量水平允许较敏感数据分析。测量要求对所测数据进行相对集中的分析。

当要求对于范围和发生频数进行编码时,假如取消了集中编码分析,就需要编码员辨别比他们所预想的更为细微的差别。另外,在编码员必须考虑做决定时,列出所有标准是困难的,因为编码人

员必须考虑到分析的主要方向。例如,编码员必须考虑测量动词时意义的相反语义关系(“赞成”对“怀疑”),以及表述所用时态的意义(过去、现在、将来)、副词修饰关系(“经常”相对于“有时”)或表达什么是可能的(经常用“可以”),以对应于强制性的话语(比如采用“必须”)。

因为这套量化编码分析有助于分析比较文字意义中的细微差别,这种定量分析研究对于分析直接引证和官方文件内容(例如公共法律条文、政府文件等)是非常有用的。在这些公共法律与规章的条文中,文字已被视为可以理解、仔细选择并转换成一定单元的准确信息。因此,在控制枪支的案例中,只有访谈书稿可被用来进行这种水平的分析。

表 1.7 通过用两种假设观点回答的态度分析,表明如何用编码表达看法集中度。即如何填写目录表格“主题、行为动作、普通术语意思”。每一种回答可以含有不止一种将被编译的叙述或记录单位,所有这些均被编码。依照“直接”和“集中度”,取值从 +3 到 -3 的范围,在每一种陈述中将被指派为动词和普通意思的术语。在这种情况下,“ + ”被指定为表示那些动词和术语以支持枪支控制的强烈程度。每一种叙述有两种取值的选择——本身动作的取值和术语取值,并且取值数量是多种多样的,且总结了回答中所有陈述的可能结果,并对每一种回答形成了一个总评分。

表 1.7　编程测量反映态度的分类

回答 1“依我之见,我赞成控制枪支,却又怀疑笼统的枪支控制是否获得成功。”

主语	动词	取值范围	术语意义取值	范围	结果
我	是	+3	控制枪支	+3	+9
我	怀疑	-2	获得成功	+3	-6
汇总					+3

回答 2“我诚恳建议政府严格控制把手枪卖给居民。”

主语	动词	赋值范围	术语意义取值	范围	结果
我	是	+3	政府严格控制枪支	+3	+9
汇总	建议、督促	+3			+9

表 1.7 的事例中,回答 1 包含两种陈述而回答 2 包含一种陈

述。第一种答案的修饰性陈述降低了强烈程度。总体上,第二种回答给出了一种较高程度的集中表述。

材料的编码

根据获得的资源和材料的版本对材料进行编码可用人工或计算机完成,这些将在内容分析的第5步加以阐述。如果已经用计算机对材料处理过,分析员应该探索获得计算机程序去编码准确度的可能性研究。在决定如何对材料编码后,分析员写出必要的说明书,表1.8列出了对于训练编码者最为基本的要求。

表1.8 编码者的最基本要求

1. 定义分析单位,包括确认他们的程序步骤。
2. 变量及取值范围描述。
3. 表述数据分类存放分析过程的要点和框架。
4. 使用和管理数据表单的指导。

预 测

在实际的编码开始之前,预测是很重要的步骤。它涉及对一部分将被分析的材料或一些别的相似材料进行试验性编码。在预测过程中,分析员检测、校对编码分类和指令,且在相同的案例分析中要重复做几次,无论用计算机还是人工进行内容分析,预测步骤都是必要的,计算机分析需要测试计算机运行状况,以确保程序按原计划准确地运行。

预测能够有助于分析员把握下列情况:①编目是否清晰和专业化,是否满足表1.2的需求。②编码指令是否充足、全面。③编码者是否适合此项工作。对编码员之间的可信度及不同编码决定的一致性进行评估(下面将讨论这一点)。一旦分析员确信文本能被可靠编码处理,就不必再进行一致性的预测评估了,编码即可开始。

当然,有了计算机程序来帮助编码,文本数据也可被编码,这样虽然解决了可靠性问题,但可能产生别的系统问题。比如,将要被编码的文本必须存入计算机磁盘(即便这也许不切实际)。另外,完成文本分析的计算机编码需要很具体的程序来控制。

例如,使用计算机通常限制分析员将词汇转化为记录单位,但这就意味着每一个将被编码的词汇必须要预先列在计算机的内存

器里类似于词典一样的清单中。然而,准备一本词典要比阐述范畴困难得多。还有,一个词汇在不同的文本中有不同的意思,计算机不能辨别这种细微的区别,而人却能够,如此也会导致计算机编码的结果可能失去有效意义。

然而,计算机有好多优点,我们不应该低估它。计算机能节省时间,且当词汇是最理想的分析单元时,允许分析大量数据。因为计算机能够比人类记忆更多的概念和定义,当编码数量巨大时计算机是很有用的。数据被多次重复使用时,多次采用计算机是很可靠的。这样,通过计算机在未来有利于管理数据的优点,将推进运用计算机分析大量研究准备的数据,并且编码分析的成本会大大降低。

可靠性检验

可靠性检验告诉分析员,采用同一测试程序进行重复试验产生相同结果的程度有多大。在文本分析中,这意味着两个或更多人把相同文本进行分类得到结论的相似性。在预测编码范围、编码种类和编码指标的过程中,分析员必须进行可靠性评估,并且这将贯穿编码的全过程。

为了检验编码的可靠性,分析员对编码员已经对相同文本编码的各种方法进行比较。如两个编码员可能对所给的十个条目各自进行编码分析,分析员通过比较他们编码的可靠性来判断不同编码结果的一致性程度。

虽然分析员一般将编码可靠性检验低于 80% 的一致性认为是不可接受的,但是也不能一概而论,可接受的可靠性比例组成最好视情况而定。低的可靠性估计不能明显揭示出分析中的错误究竟出在分类过程还是出在编码当中。所以,在预测中,分析员识别有差别的编码的主要内容,且了解其原因是非常必要的。如果假定编码员是能够胜任的,出现低的可靠性估计就意味着编码员需要就一致性差异进行细致的解释,而不是去训练编码员去了解分类过程。

解决问题的途径之一是对已经被进行了可靠性编码的数据与还没有办法编码的数据进行比较,这样做会告诉分析员错误是集中存在几个分类(范围)内,还是贯穿于全部系统当中,如果存在于全部系统当中,分析员应该认真考虑整体设计,包括使用文本分析的决定,如果只是几个小范围内存在问题,那么修改这些范围(或编码指令)或许会解决问题。

分析和解释结果

内容分析的主要目标是分析具有固定格式的信息，这些信息已被转换成可用的格式，这一点将在第6章中论述①。

- 总体描述编码数据。
- 发现编码数据类型及数据内在的关系。
- 确定测试数据类型及其关系的假设是否成立。
- 将从其他方法和条件对内容分析的有效性的评估联系起来，以获得的综合全面的数据结果进行总评。

这些任务或完成这些任务的分析技巧对于文本分析都不是独一无二的，根据编码设计，分析人员可以辅之以使用大量的统计方法。

概括数据和检测数据类型

归纳数据最普通的方法是通过观察数据在文本中出现的频数。绝对频数是指陈述或论点在事例中出现的次数，相对频数可通过绝对频数与总数的百分比表示。分析员可以把一种文本的频数与所有文本中的平均频数做比较，或许他们能注意到一段时间内频数的变化。

例如，在美国住房与城市发展局的评价体系中，GAO分析员已经把来自两个办公室的38份评估报告的观点做了分类。分析人员对于这些报告中的每一个主题都逐一进行了统计。他们采用了绝对频数法来计量报告中的观点，统计结果就形成了表1.9所示的状态。在统计中分析人员发现，38份文件中的20份没有就住房与城市规划方面的关系陈述观点，另有16份文件没有署名调查抽样的来源。

文本分析的另一种方法是通过列表表明变量的出现来检验变量间的关系。例如，表1.9表示各种报告间和形成报告的评估单位之间的联系，从这些信息中，GAO分析人员确认来自两个办公室的报告所引用的观点几乎出自于同一文献来源。

列表并不需限定为二维或三维度的变量分析，多变的方法可用

① 本书没有收入该书第6章的内容。——编者注

表 1.9　从美国房屋与城市发展局的一个评估项目中列举的观点频数统计表(示意图)

Y	Z	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
A																																	
B																																	
C	15														2		1	2	1										3	2			14
D	3																1																2
E																																	
F	7						4																								1		3
G	3					2																											1
H	30															2	2	1						4		1	3	1					20
I																																	
J	4	1						1		3																							
K																																	
L	1											1																					
M	2												2																				
N																																	
O	0																																
P	1											1																					
Q																																	
R																																	
S	8																		1	1		2			1	3							
T	38	1				2	4	1		3		2	2		2	2		2	2	1		2			1	4		1	3	2	1		20

资料来源:U. S. General Accounting Office HUD's Evaluation System — An Assessment PAD-78-44(1978). p. 7　收入本书时有改动。

表1.9注:

1—房屋交易由政府统筹;2—加强审计;3—银行提供现金贷款;4—较低生活水平者的房屋补贴资助;5—在政府房屋规划中添加住户财产评估;6—对于8项条款提供房屋援助;7—按照8项条款进行住宅街道规划;8—建立8项条款要求的财政援助;9—修改8项条款中的财政规定;10—对于房屋规划的行政干预;11—选择235商业服务模式条款;12—在FHA住房项目上使用第八条款;13—房屋租赁中的市场公平限制;14—现代化的公共房屋建设;15—可以替代FHA计划的项目;16—房地产交易条款修订;17—国会对于房屋财产立法的保障;18—对于FHA项目评估;19—房地产年度增长评估;20—房地产增值的补偿金发放;21—房屋与经济增长的关系评估;22—CDBG项目的作用;23—房屋建筑工地监督;24—住房等级与城市规划;25—CDBG项目的应用;26—CDBG项目工程概算;27—对CDBG项目的财政资助;28—CDBG项目准则;29—CDBG项目与住房;30—有关住房立法701条款的效果和影响;31—住房法律中312节法律效果的评估;32—没有其他观点。

A—分析单元的影响;B—住房贷款保证;C—调查;D—其他;E—房屋等级;F—研究;G—另外;H—评估总体影响;I—专门的分析单元;J—房屋援助;K—房地产担保;L—调查;M—其他;N—街区规划;O—调查;P—其他;S—专项研究合计;T—总计。

Y:内容分析区域划分和项目分析研究单元

Z:统计研究的主题种类。

来分析复杂结构,在数据中发现类型和关系的别的方法还包括偶然性分析、聚类分析和因素分析。

有效性评价

无论使用什么样的方法,最终并且最重要的目标是,通过把现有的数据与被认为合理的有效性结果联系起来以评价结果的有效性。有效性某种程度上是对用方法测量将要打算测量的东西的可测出程度精度的预估计。对于通过内容分析所做的有效性推论,可靠性和大量事实是必需的,但还没有构成采用内容分析进行科学推理的足够条件。另外,分析员必须通过采取其他佐证方式证实内容分析的结果,这需要通过对于所分析的现象与数据进行关联性的分析处理来达到这个目的。

罗莫洛有关和平志愿队伍在非洲科罗斯路得地区一个和平合作的经历描述案例的报告,提供了有效性评价的一个例子。他报告的文本分析假设能把成功的志愿者与不成功的志愿者区分开来,假定不成功的志愿者表现出疏远他们的经历。对同样的志愿者,罗莫洛把他的结果与管理者等级评估结果相比较,发现两者之间的相关性。归纳出他自己的分析将产生一种有效志愿援助的成功测量。

其他的同等程度的测量可用来证实罗莫洛的发现,对与志愿队伍合作工作的非洲人的调查就是一例。通过测量食品生产的增加

与婴儿死亡率的下降相对比，对每一个志愿者所在村庄与没有志愿者所在村子相比较，这种可以大样本量的使用和普遍接受的确证测量方式减少了由于主观评价误导发生错误的可能性。

撰写报告

在撰写 GAO 的任何报告时，分析员应该就他们工作的范围、性质给读者阐明报告中哪些是分析人员发现的证据、哪些是分析人员在研究中所依据的参考框架，以便让读者读后有一个清楚的思路：内容分析已经做了什么、为什么做、结果为什么能为结论和归纳提供一个稳固的基本证据。表 1.10 描绘出了写作报告的要点，这种信息是当分析员使用文本分析时应该主张强调的。

表 1.10　一种文本分析研究的最少资料

1. 研究对象及其数据选择、方法和研究设计的总体思路。
2. 数据选择、方法和设计的充分理由和原则。
3. 分析过程描述（以便别的研究者能够重复进行），尤其包括下列描述：
   - 调查抽样计划
   - 分析单元选取
   - 编码规则
   - 结果的可靠性分析
   - 数据处理、分析方法过程
   - 局部方法和全部方法生效以后预期产生的效果
4. 研究结果和统计意义总结。

内容分析的结果应该是相当严格的，足以承受可靠性检查。表 1.10 提到过的条款所阐述的信息或许包含在报告或附录的主要部分里，或者仅留在报告的正文当中。

另外，或许还有另一种情况，那就是内容分析的结果可被完好记录下来，让那些挑剔的读者评价报告结果可靠程度究竟有多少。

本章小结

内容分析是把一些非结构化的、无序的文本转化为可以用来分析的格式化文本的一系列方法。在本章中，我们已经描述了这些方

法,现概括如下:

1. 基于分析项目文本的客观目标决定使用文本内容分析,这些被用来分析的文本是可得到的、可获取的材料,并且能够用来进行所需要的分类比较。

2. 决定包含在内容分析中的材料及涉及样本抽样的材料。

3. 选择语篇的分析单元和编码记录单元。

4. 开发编码目录内容、量化分析水平和编码原则。

5. 预测编码目录内容,然后通过手工或计算机进行文本的编码。

6. 在测试和完成编码的过程中,随时检验其可靠性。

7. 分析和解释编码数据,评估内容分析结果的有效性。

第2章

GAO的分析人员为何要使用内容分析法研究政策

本章中，通过提出采用或不采用内容分析的理由，概括出前面介绍的总体结论。讨论内容分析的优缺点，然后给出在GAO工作中内容分析方法简要的应用情况。

内容分析能做什么

所有研究者要想系统地分析文本，都必然会考虑到分析文本的内容。因为，内容分析是从已有的数据源中提取获悉新的观点的分析方法，因此，它是一种几乎应用在所有项目中都非常有潜力的分析方法。

能提供严谨的测量

存在于文本或报告中的文本内容分析最初会使人产生一种不宜察觉的、难以引起注意的感觉，从而对这种方法也反应漠然。这正如在采用非文本分析调查所反映出的普遍存在于实证方法中的一个常见的问题是：不采取内容分析时，分析人员和他们所分析对象间的相互作用会引起主观上反映出其他贴标签的评价情况，而不是他们自己较为“符合自然本质”的主观推理方式。且这种主观判断会把偏见带进结果中。另外，对于个人面对面地进行刨根问底式的调查被认为是严重侵犯人权的行为，因为这些直面的问题难免会触及被访谈人的个人隐私，一般会被习惯地从分析中剔除掉。而内容分析的优势在于，采用中性的存在于文件中的量化表述，这样的

文本分析有可能避免以上两方面问题。

能同时进行大量文本的分析

因为内容分析清晰的编码指令、精确的分类和一定的可靠性检查,这使得有一定文化水平且接受过内容分析方法培训的任何人都能对文本进行分析,所以大量的文本借助于内容分析被分析。同时,大量的文本允许两个或更多的编码员对同类数据在不同的地点进行处理,如在总部办公室和分部办公室。

内容分析法有助于分析员了解真实的事件范围

内容分析法有助于分析员更多地了解他们正在调研的内容和观点。这一优点起源于两种特性:其一,内容分析本质上是系统分析,其二,由于编码方法是定量和客观的,它的可靠性分析和分类编码的原则是客观有用和非常严谨的。

内容分析能验证其他方法是否有效

在第 1 章中,讨论了通过采用别的方法所得结论确证进行内容分析的有效性。这种有效性也可以沿着相反的方向拓展,也就是说,来自于内容分析的材料可被用来验证别的方法得出结论的有效性。例如,调查数据和委托研究之间,调研员如何用"多种测试"来确证他们文本分析的可行性和可靠程度。

运用内容分析过程中应防止的"陷阱"

以上我们竭力展示内容分析的优势,大量现有的材料可能诱使分析人员掉进一种漫无目的、代价昂贵的"沉溺于编码的迷宫阵"中。另外,将文件进行编码统一量化成为一系列数据,也可能导致一些很重要的数据被埋没在大量平庸的常规统计数据中,无法突出其中的重要性和利益所在。将文件信息数量化表示也可能产生重要的有意义的数据,但也容易使研究限于大量琐碎的、意义不太大的描述性统计当中,这样的操作最终很可能会产生分析准确但结果无意义或价值不高的东西。

费用高昂、耗时持久的内容分析

内容分析的特点是成本相对高、工作费时,对内容分析的使用

者访谈以及查阅内容分析方法总结报告所得出的共识就是要在以下三方面采取措施，以降低内容分析的费用：

- 采取程序化的编码分类不一定十分可靠，并且由于其中操作步骤重复会耗时过长。花在这些程序化编码操作上的时间可能会从几天到两三个月不等。
- 如果编码人员涉及较多数据分析，并且必须要用手工分析处理数据时，就必须对从事项目的编码人员进行培训。如果分类计划，文本分析特别耗时，并且编码复杂难以简单分类，这时，若要采用内容分析作为研究方法时，就须计划好人工编码、培训管理编码员所需要花费的大量时间和费用。
- 如果需要分类编码的工作量巨大且记录单元较小，内容分析工作量琐碎（例如，需要分析一些单词或题目），且分析的单元文本容量大（例如，篇幅较长的报告）时，由于编码必须有系统，因此分析过程可能冗长和乏味。使用计算机代替人工编码过程可以提高效率，但其前提条件是需要为用计算机处理编码问题预先编制一系列的标准编码表列，以便使文本计算机化有所参照。

内容分析的信度、效度问题的测度分析

信度和效度是相互依赖的概念。从总体上看，准确定义目录能引起高可靠性和统计的重要性，但会减低分析的实际语义，而模糊分析又会降低分析的精确性，所以，必须采用折中的办法。对客观要求和重复分析的需要可能迫使分析员对他们感兴趣的东西先行编码以取代用机器编码，这样预示着保持分析的有效性。重新定义分类来增加其有效性可能会意味着导致分析变量之间关系的不协调。所以，面对这种进退两难的境况，必须在编码分类拓展后进行有效性评估。

程序评估的潜在用途

内容分析为何应用在 GAO 工作被认为是出于以下三个因素：项目评估要求、被分析材料的特征要求，以及分析方法的要求，我们给出集中在三个分类评估目标上应用案例的主要情况来表明文本分析是如何被实际应用的。

定义分析项目的目标

采取项目评估的客体需要首先定义项目的预期目标。为此,项目分析人员需要广泛收集分析资料,从有关的计划、政策文件,或访谈的会议记录,收集关于立法过程中的手稿或者手写的、打印的信息。使用内容分析,分析员对材料的观点可以客观和系统。除了为辨别程序目录提供给分析员的结构化公式外,这种技巧还可推动决定那些目标是否与立法的目的有矛盾,如果允许的话,将官方的文件与有关会议文件作对照。

描述项目的积极性

项目评估可有一个对项目积极性的客观描述,为了实现这一目标,分析员可研究形势,参加代理会或会见项目分析的设计者,通过这几种方法收集到信息后编成文本,紧接着采用内容分析来研究编码这些文本。

从这种分析中,必须首先对所收集到的文本进行大量简要、客观的文献综述,这是生成更多复杂编码设计的基础。内容分析可以被用于对一个主题事例贯穿于某一时间段活动趋势分析,可以用于整合若干有项目的目标活动,可以对于服务性质的项目进行信息反馈规律的调查分析,可以进行商业服务满意程度的调查分析,还可以为若干种类的信息比较和一些访谈提供一些行为的准则。

评估项目的结果

项目评估很可能能够查明项目的结果是否客观。或许作为分析专家,需要对于一个政策执行的实际结果有充分的把握。在此形势下,分析员可通过研究早期的评估报告或通过调查参与程序者收集信息。在调查中,开放式的问题可能有益于从那些不宜被格式化确定的论点、观念、态度的收集中获得所需要的信息。使用内容分析可以分析开放式的问题,分析人员不需要把自己的观念强加于调查回答的结果之上。对开放式的调查数据使用内容分析法,这样分析员能检查一段时间内项目完成的趋势及将实际情况与项目活动的目标相比较。

或许还有另一种评估方式,即通过对于项目所涉及人员情况分析来评估项目发展的趋势,这些被研究人员被按照所在的地理位置、年龄、收入和长相划分成不同的种类,以便于进行联系。

本章小结

我们希望给已经看过前面文章内容的读者就内容分析方法留下一个真实的感觉，即内容分析法有其优点也有其缺点。方法总是有其局限性的。假如没有明确、清楚的目标，内容分析要得出准确的结果，这几乎是不可能的事。内容分析方法在形成分类目录时是很费时、费钱的，这些分类目录可以用于可靠地进行编码统计。

提出编码原则、培训和检查考核编码人员都是很耗时的。另外，复杂的编码项目或许会产生极有趣的发现，或许产生可靠性低的结果。成功与否取决于内容分析之前预先确定的有关编码判断的重要准则。所以，如果内容分析要经过承受真实性检查，就需要严格的可靠性和有效性检查，或辅之以其他资料分析的佐证。另外，分析结果也取决于被分析文件中所包含的多元信息数量的复杂程度，假如这些信息是不可靠的或无效的信息，即使最严格的内容分析也只会得出无任何价值的结论。

然而，内容分析将进一步应用在几乎所有的政策项目研究中，还被应用在政策过程的任何一阶段。在开始阶段内容分析确实有助于用来帮助分析员了解项目进展的实质性过程。相对于过去惯用的收集存在于数据资源中的项目信息，内容分析是一种相对完美的方法。它不需要收集数据，这就意味着节省时间和金钱。在文本中我们所讨论的应用的可能性是无限制的。确切地说，我们本打算展示内容分析的多样性，然而内容分析能被应用的数量和种类，以及能被帮助回答解决的问题实际上却是十分有限的。这就是内容分析方法。

第 3 章[①]

对政策内容进行量化分析的研究路径

在对决策进行研究的过程中，我们至少可以从文献中获得两类有用的信息。不论是新闻工作者对于事件的描述，还是相关的编年史，都为我们分析影响决策的环境因素提供了大量真实的素材。因为分析时选取的都是文献中引人注意的内容，所以并不具体规定必须使用什么样的研究方法。对决策进行研究的学生，可能也想了解决策者个人的信息以及反映决策情境中决策者之间的相互作用的信息，并以此为根据来验证假设。这种研究设计要求对数据进行系统的收集和整理，所以，必须使用固定的研究方法——内容分析来解决问题。

作为众多研究方法中的一种，内容分析的运用范围很广。文学研究使用该方法来解决争论不休的作品归属问题。宣传分析(propaganda analysis)则很大程度上扩展了内容分析方法研究的范围：社会学家们在战争间歇期内使用内容分析的方法来探究宣传的技巧；第二次世界大战期间，美国通过对敌方的宣传进行不定期的内容分析而获得了一些军事情报。第二次世界大战以后，很多人尝试着运用内容分析的方法来研究国际政治中使用符号的变迁状况。最近，内容分析成为心理语言学家使用的主要工具，他们用它来分析信息与信息提供者的个性特征之间的关系。

多温·卡特赖特(Dorwin Cartwright)简单明确地说明了在国际

---

① 本章内容引自由迈克尔·哈斯(Michael Haas)等起草于 20 世纪 60 年代早期(大约应该在 1963 年之前)的一篇有关国际关系政策的研究论文。该论文的版权和著作权已经全部开放，可以自由引用。原文由江雪梅、李钢编译，李钢对译文进行了统校。

冲突和整合中使用内容分析这一方法的主要原因：

> 尽管社会和政治冲突常常是由于经济利益的冲突和力量的强弱不均而导致的，但是，如果我们不研究冲突各方在交往过程中所使用的语言文字，不研究构成大部分谈判过程的调解程序，我们就不能完全理解这些社会和政治冲突。

在这种境况中，内容分析的研究方法得到了广泛的运用。我们可以使用它来研究冲突性目标或者意识形态的内容；我们还可以使用它在不同层面上进行分析，以系统地取得涉及某些人的认知、评价或者情感状态的信息，这些人的话语权力与他们的国家大政息息相关，这些人的个人决定往往就是国家的决策。

简单地说，大多数的研究调查都可以使用内容分析的方法。在进行社会科学研究时，人们越来越倾向于质询一些具有普遍意义的研究问题，虽然这样做已经导致设计出来的研究方案越来越显现为强词夺理，但是，就内容分析这种研究方法来说，仍然不是简单地用一个"好"字就可以概括的。

设计研究方案

内容分析研究通常包括下面几个阶段：首先，提出研究问题、研究理论和研究假设，进行抽样，确定类别。其次，阅读文献，对文献进行编码，使用特制表格对相关内容进行精简处理；文献编码完成以后，评定各类别下每一项的权重指数，并依此确定每一项的频数。最后，使用恰当的理论论述研究结果。

陈述问题

研究问题的确定总是先于研究方法的选择，接下来是对假设进行具体的描述，并对概念加以限定，从而研究才能够进入操作阶段。

研究国际冲突的方式一般是，分析处于危机情境里的决策者，确定该个体对于事实的想象和理解、他的心理状态及其思维链（means-end chain）的结构和模式。有一种观点认为，关注情境中的情感因素而不关注其战略意义的人，易于做出开战的决定。另一种观点则认为，如果消极影响在决策者的表述中处于不断增加的状态，那么决策者就倾向于寻求一种"调节行动"，比如"增强对抗"，来缓解其言语导致的紧张。总之，研究假设一旦确立，研究者的任

务就是思考相关的抽样程序、确定文本类别,并对收集到的数据进行编码。

选取样本

将研究问题控制在可操作范围内的方法就是,研究者从全部可用数据中进行抽样。抽样的过程可以是抽取一个案例和数个国家进行研究,进而选择一些人物和文献,分析概括而得出内容分析所需要的数据。

选择供研究的案例和国家时,我们必须同时考虑理论因素和实践因素两个方面。国际冲突和整合研究中,将"决策者在危机情境里的行为"确定为研究内容。研究者之所以选择了1914年欧洲危机作为研究案例,根本的原因在于可以收集到这一时期里非常完整的可用文献;同时还因为这场战争的爆发,既不是人们有意识地策划的,也不是战争参与各方预先所期望发生的。虽然1914年危机涉及了很多国家,但是研究者只选择了其中的五个主要国家,对这五个国家的决策者的各个方面进行内容分析,因为这五个国家的可用文献相对来说比较完整并且具有权威性。这五个国家是英国、法国、德国、奥匈帝国和俄国。这一时期,意大利和塞尔维亚的文献,除了保存在罗马和贝尔格莱德的以外,其他的都不完整,而且大部分不可用。

为了研究一个国家的外交政策和决策过程,研究者必须明确谁是国家的决策者,即那个经过授权而与国家命运紧密相连、对国家负责并追求国际地位的人。为了抽样的方便,国际系统中的"执政国家(behaving state)"一般限定为包括国家的名誉领导、总理大臣和外交部长在内。然而,这个系统只是一个基本的模式,在不同的研究项目中需要有细节上的调整。1914年危机的研究和20世纪中期中苏联盟的研究中,研究者选择了下面这些人作为各自国家的代表①。

---

① 从1914年研究中,可以得出法国、德国和奥匈帝国的决策者的名单。对比研究根据1949年以后中华人民共和国的政府结构,除了国家主席、政府总理和外交部长以外,还有必要罗列出其他对政策的制定起着明显的权威作用的决策者,也有必要收集反映官方政策的那些资料。这些资料包括共产党主席毛泽东的讲话,共产党理论发言人郭沫若的言论,以及官方报纸和杂志上发表的某些不署名的社论。

1914 年的决策者

德国:

威廉二世(Wilhelm II)——德国皇帝

西奥博尔德·冯·杰戈欧(Theobald von Jagow)——德国外交部长

阿尔弗雷德·冯·齐默尔曼(Alfred von Zimmermann)——德国外事办公室副主任

英国:

乔治五世(Gorge V)——英国国王

赫尔伯特·阿斯奎思(Herbert Asquith)——英国总理大臣

爱德华·格雷爵士(Sir Edward Grey)——英国外交部长

亚瑟·尼科尔森爵士(Sir Arthur Nicolson)——英国外交部常务副部长

中苏关系研究的决策者

中国:

毛泽东——中国共产党主席

刘少奇——中华人民共和国主席

周恩来——中国国务院总理

陈毅——中国外交部长

前苏联:

利奥尼德·勃列日涅夫(Leonid Brezhnev)——前苏联总书记

尼克塔·赫鲁晓夫(Nikita Khrushchev)——前苏联总理

安德鲁·格罗米科(Andrei Gromyko)——前苏联外交部长

抽样的方法取决于所研究问题的区域特征。样本的基本来源是决策者亲笔书写的文献资料,次级来源是包括决策者的发言或大家公认的权威对文献做的说明。研究实践证明,下述原则有助于资料的选择:

1. 可以接受的基本资料:

(1)政府之间的官方交流。

(2)外交指令、军事报告、部委备忘录、内阁会议或者委员会会议记录、议会争议、论文集或者日记(如果真是由决策者本人撰写的)。

(3)决策者亲自起草的,或者授意于某个决策者而起草的文献中的部分段落。

2. 可以接受的次级资料:

根据与主要决策者的会面而撰写的大使报告或者新闻报道

(前提是:大使或者新闻作者仅仅引用决策者的话语或者解释,而不加入有自己观点的话)。

有一些关于 1914 年危机的文献资料,特别是所谓的"×皮书(color books)"都是经过篡改了的,一点也不可用。另外,当时的文件,虽然没有包括在文件汇编中,也是可用的;日记的可用性应当优先于论文集,因为论文集可能是后来编纂的。研究者在编码之前就应当将所用的资料确定下来。

这个阶段中,调查者依据自身的研究兴趣和数据处理的规则,对即将进行内容分析的数据进行筛选。下一个步骤就是构建数据类别,对数据进行分门别类的处理,最终确定数据分类体系。

建构分类指标

确定数据种类的目的是在已定概念的基础上对数据进行分类,以便研究者测量每一类别中变量的频数。

确定数据类别时,必须先对之进行诸如效度、信度和客观性之类的标准评价。效度指的是测量工具对被测物体进行测量时实际可以测量到的精确范围。如果我们使用两种不同的工具对同一现象进行测量时可以得到一样的结果的话,那么这种分类方式就是有效的。为了保证分类具有较高的信度,我们在分类过程中还必须保证,一旦重复同样的研究过程,所得的结果仍然属于规定的信度范围。信度有两种,一种是编码者之间的信度(intercoder reliability),指的是不同编码者在某一时间进行测量时,所得结果的一致性程度;另一种是编码者信度(intracoder reliability),指的是某一编码者在不同时间里重复测量时,所得结果的稳定性程度。如果根据某种分类方法,可以得到无偏见的数据,也就是说收集到的数据与数据收集者的个性特征毫无关联,那么,这种分类就具有客观性。因此,效度、信度和客观性是建构分类指标时必须考虑的因素,反过来,分类指标的构建也决定了编码以及测量的有效性。

可以进行分类,进而进行分析研究的概念,大致分为两大类:一类是情感的或者可评估的,另一类是认知的,或者更确切一点说,是

指示性的(denotative)①。表达情感的或可评估的概念，显示了措辞(expression)中的个人态度部分，而表达认知的概念主要是描述了一种情境或者一种行为。指示性概念可能包括政策情境、冲突的解决以及能力；有关情感的概念可能包括力量、友谊、敌意、满意和挫折。其他的分类标准同样也适用于这种研究。为了帮助读者更加清晰地理解建构分类指标的意义，我们可以通过"1914年危机研究"和"中苏联盟研究"这两个例子，来说明上述几种分类方式。

(1)政策情境。揭示了思维链(means-end chain)上一个个的行动及行为习惯，不论这些行动或行为习惯是发生在过去的还是发生在当前的②。其内容包括目标、目的、偏好、选择和手段——通过这些手段，国家可以达成或者期望达成这些目标。国家的目标可以是长期的、终极的，也可以是具体的、短期的，国家可以自由选择达成这些目标的手段。例证：

> 总理大臣赫尔伯特·阿斯奎思告知议会(1914年7月29日)："正如王室所知，奥地利昨天向塞尔维亚正式宣战。"
>
> 赫鲁晓夫总理就美国有关U-2战斗机的政策发表评论(1960年5月20日)："……美国总统保证……暂时停飞，1961年1月解禁。"

(2)冲突解决。包括两个方面：①评析解决模式的可行性；②预测冲突解决可能带来的后果。它揭示了当事一方对冲突解决的特性及相关后果的看法。例证：

> 亚瑟·尼科尔森爵士针对即将发生的冲突提出了大致的解决方式(1914年7月28日)："……在那个事件中(奥匈帝国侵占塞尔维亚)，所有和平解决的希望都是徒劳的。"
>
> 德国皇帝威廉二世在第一次世界大战爆发前夕预测(1914年7月30日)："……如果我们将要流血到死，那么，英国至少也得要失去印度！"

---

① "认知的"、"情感的"和"可评价的"三个术语在使用时存在着一定的差别。关于这些术语的使用问题，社会科学家们仍然争执不休。本书中，这些术语是在操作的层面上使用，而不是定义性的使用，因此，这些术语都有待于进一步的精确。

② 政策情境的某些方面可以被分离、测量并做出独立的分析，这些方面是：现状的变化，即政策可以改变国家间现存关系的程度；特性，即焦点，政策所反映的特定的即刻的方面；一般影响，即政策情境表达中所体现出来的一般情感性内容；最后一方面是暴力，即政策作为一种解决问题的方式所表现出的暴力性程度。

(3)能力。包括关于经济财富、军事武器、军队等总体数目的积极的、非暴力的观点,这种划分是对一国的国力进行定量分析,而不是定性分析。当某些项目构成了政策情境时,就可以使用这种划分方法。例证:

法国外交大臣评价比利时的国力时说(1914 年 7 月 28 日):"……回想 1912 年、1911 年和 1910 年……对照 1913 年,可以看出比利时军队的总数达到了 10 万人。"

赫鲁晓夫总理告诉(前)苏联总书记(1960 年 1 月 15 日):"苏联朝向月球发射的那枚火箭是太阳系里第一颗人造星星。"

(4)国力。这一类别包括对国际系统中相关各方的相对优势和劣势进行评价的陈述。这种评价是定性的,不是量化的评价。例证:

在一份送交德国政府的外交照会(1914 年 6 月 27 日)中,奥匈帝国政府是这样评价土耳其的国力的:"土耳其……是俄国和巴尔干半岛诸国之间的一个强有力的平衡力。"

李富春声明(1960 年 1 月 1 日):"毫无疑问,我们有能力实现连续跨越性的经济发展。"

(5)友谊。是各参与方之间相互表达出的认可感和积极行为。它显示了各国之间相互支持、共同努力,为实现目标而进行的积极联络。例证:

弗朗兹·约瑟夫国王写信给德国皇帝威廉二世(1914 年 7 月 2 日):"谨向您表达我诚挚的谢意……"

《北京周报》一篇关于苏哈托(Sukarno)总理的社论报道:"他在访问中国期间发表了关于中国和印度尼西亚两国之间友好关系的热情洋溢的话语,给我国人民留下了深刻印象。"

(6)敌意。一方向另一方表达出的憎恨和负面的感觉①。它表明一方蓄意阻挠另一方实现目标的基本态度。例证:

---

① "敌意"这个概念包括三个方面:a. 如上所述的情感因素;b. 如"德国侵占了比利时"中所描述的行动因素;以及 c. 在"我们将消灭你"中表达的威胁因素。本书所说的"敌意"这个概念指的是一般意义上的敌意,为了操作,必须对之进行精确的区分。

德国皇帝威廉二世这样描写英国的政策(1914年7月30日):"……她(指英国)使得我们的政治和经济走向衰退。"

《人民日报》社论宣称:"中华人民共和国坚决反对日美军事联盟。"

(7)满意。成功和满足的感觉以及某项行动被接受后产生的愉悦的感觉。它关注的是人们对自己在系统中的地位表现出的积极的感觉。例证:

奥地利在送交德国的一份文件中,记录了保加利亚政策中的一个令人满意的变化(1914年6月27日):"但是最重要的是,保加利亚已经从俄国施下的魔咒中解脱并清醒过来了。"

赫鲁晓夫总理声称(1960年5月19日):"我很高兴地说,我们的共同愿望是继续努力多向度地发展我们之间的关系。"

(8)挫折感。即一方表达出的失败和失望的感觉,其结果在于妨碍政策的贯彻实施。它表达了该方的焦虑和不适,强调的是其对自身和自身在系统中地位的消极感受。例证:

奥地利宣布(1914年6月27日):"……现在的局势……从三方联盟(the Triple Alliance)的视角来看,不能说对各方都有利。"

编　码

在"国际冲突和整合研究"中,我们选取"主题"——而不是单词——作为分析的基本记录单位。编码时,我们将段落和句子分解成"基本主题(atomic themes)"或者"概念单元(unit-perceptions)",意思是,这些"基本主题"或"概念单元"不能进一步分解,一旦分解便没有任何意义。下面给出一些"主题"的例子,这些例子都是从文献研究中筛选出来的:

1."我认为,俄国外交大臣建议与奥地利政府友好地交换意见……"

2."塞尔维亚人在市区和郊区进行的宣传最终导致了萨拉热窝危机。"

概念单元,通过三个独立分离的因子水平(actor-levels)表示出来,这三个因子与已知概念中描述的行动或结果有关联。我们进行编码操作——进而测量——的基本"单元"由四个部分组成:①感

知者(perceiver);②被感知者;③行动指向的目标或者影响的对象;④描述性联结(descriptive-connective),或者是行动的本质或各因子间相互作用的本质。

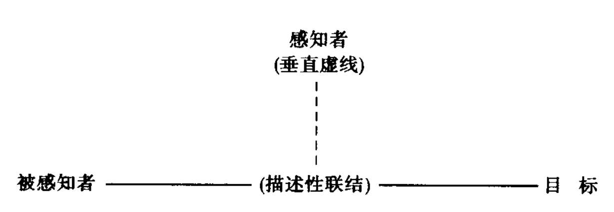


图 3.1 编码操作的基本单元

从图 3.1 中可以看出这些成分之间的系统关系。箭头代表行动或影响的方向,垂直的虚线显示了刺激和感觉。描述性联结是概念单元中可以进行测量的部分,它显示了变量的本质特性。

编码以及确定概念单元时应当遵循的基本原则是,每一个设定的主题(assertion)都只能有一个感知者、一个被感知者、一个目标和一个描述性联结。然而,在既定的概念单元中,比如"战争进行得如火如荼",我们并不能明确地指出被感知者和目标。此时,将被感知者和目标注明为"一般"或"不确定"更为妥当。如果不考虑这些限定因素,那么,所有的概念单元都包含有一个行为或影响的明确感知者,同时也包含有一个描述性联结,因为,必须有一个明确的影响因素,人们才会感觉到,而且主题也必须包含确定的内容。

除了每一个因子都可以发生变化之外,利用新的类别维度也可以显示出既定的概念单元的范围。例如,这一段文本,"奥地利的意图,同时也是德国的意图,就是彻底消灭塞尔维亚",就应当作为两个独立的政策情境来编码:

1."奥地利的意图……是彻底消灭塞尔维亚。"

2."德国的意图……是彻底消灭塞尔维亚。"

上述例子说明了一个原则:对于感觉来说,如果感知者、被感知者或者目标发生了变化,那么就会产生新的概念单元。

下述"中止原则(cut-off rules)"揭示了怎样判断一个主题结束、新的主题同时开始的方法:

1. 一个主题仅存在于一个段落内。

2. 在下面的任何一种情况下,都产生了新的概念(或主题):

(1)"感知者"发生变化。

(2)"被感知者"发生变化。

(3)“感知的目标”发生变化。

(4)种类发生变化。

可以使用多种形式的编码表记录和保存内容分析的数据。“国家冲突和整合研究”中使用的编码表,概括了从文献中提炼的信息,并确定了相关的因子。为了避免情境的缺失,编码的过程中再现了主题提炼的文本。每一个必须逐字再现的主题,都归入“命题陈述(capsule statement)”一栏,并标明“元音省略”、“插入”或“引用”。表3.1是其中的一种编码表(这是一种直接针对文本设计的开放式的草表)。

表3.1 编码表示例

指示性概念:一、______ 二、______ 三、______

文献作者:赫鲁晓夫 被引用者:N/A 文献接受者:新闻发布会

对话者:N/A 资料来源:《真理报》1960年5月19日

“感知者”:作者X 接受者___ 被引用者___ 对话者___ 其他___

“被感知者”:作者X 接受者___ 被引用者___ 对话者___ 其他___

感知的“目标”:美国

本概念的由来:事件A,其历史___现状___发展趋势____________

〔文本〕:

……当然,艾森豪威尔总统可以决定是否派出美国的飞机,而他们的飞机能否飞越我国领空则是另外的问题了。这由我们来决定,而且理所当然地由我们来决定。我们要把飞机打下来,我们会根据飞机碎片的来源分门别类地进行管理,我们会根据飞机的制造原理和实际组成来管理这些飞机碎片。

命题陈述:

一、1. 这一点前(美国能否飞越前苏联的领空)由我们决定。

2. 我们(前苏联)要把(美国)飞机打下来。

3. ……我们(前苏联)会根据(美国)飞机碎片的来源分门别类地进行管理……

4. ……我们(前苏联)会根据飞机的制造原理和实际组成来管理这些飞机碎片。

二、……

三、……

= =

一、有关“政策情境”的概念

二、有关“能力”的概念

三、有关“冲突解决”的概念

编码者首先通读全部文献,然后仔细阅读每一个段落;如果有的文献不分段落的话,那么可以将两英寸宽的内容(two column-inches)视作为一个段落。一旦编码者在文献中确定了一个概念,他就要相对地填写编码表,方式如下:

1. 标题是"指示性概念(denotative perceptions)",标题下面第一行叫做"该主题的作者(author of statement)",当然这个作者就是这段文字的作者。"主题接收者(recipient of statement)"是这段文字呈送的人或国家。"被引用者(person quoted)"指的是文献作者所引用或评析的文章的作者。如果被引用者的这段话是以口头语言的方式传递给对方的,这个"对方"就是"对话者(interlocutor)"。

2. 第三行是"资料来源(source)",表明文献的出处和文献产生的日期或者发表的日期。

3. 第四行和第五行,是有关谈话者和对话者的信息。在赫鲁晓夫的声明"俄国的政策是和平共存"中,"感知者"和"被感知者"都是作者——前苏联,其代表是决策者赫鲁晓夫——因为,是前苏联在感知她自己制定的政策。如果赫鲁晓夫在联合国大会上发表演说,说他认为联合国故意延缓惩罚侵略者政策的实施,那么,"感知者"就是前苏联领导人(作者),而"被感知者"则是接受者(联合国)——是前苏联在感知其他人的政策。

4. 接下来的一行叫"感知的目标",就是政策或行动指向的对象。若在上例中,联合国的政策是针对刚果而制定的,刚果就是目标。

5. 在标有"文本"的栏目内,编码者必须将抽象出主题的段落全文抄上,这样做的目的是为其他参考者提供语言资料环境。

6. 在标有"命题陈述"的栏目内,编码者应当写下的是研究确定的基本主题,填写时倾向于直接引用,仅仅删去与该主题无关的字句。

"信度"与"效度"

编码数据的信度,是经过训练的编码者熟悉的、经常使用的概念之一。数据的效度,数据对于结果的概括能力,都与信度直接相关。因此,在研究的各个阶段中,我们都应该进行信度检验,评价编码者之间的一致性程度以及编码者的前后一致性程度。在培训中,向编码者出示一些文字段落的样本,让他们编码,藉此对之进行信

度检验。一般是在研究开始及结束时，进行信度检验，检验的方法是对研究中已经编过码的材料进行重复编码。

信度的计算公式有很多，不同情况下可以使用不同的公式。我们可以从其中选择与研究相符合的信度计算公式。当研究仅仅包含两个编码因子时，计算公式最简单：

$$R = \frac{2C_{1,2}}{C_1 + C_2}$$

所有编码者都一致同意的编码种类的数量除以全部的编码总数。

如果编码者一致性（intercoder agreement）程度低于可接受水平，那该怎么办？研究者可以继续培训编码者，也可以重新定义或压缩分类标准，或者干脆调整分类的数目，以便于单个编码者承担责任。编码者的一致性信度较低的原因，常常在于编码者们对分类标准的理解不统一，当编码者所依据的分类标准表述不清或冗长啰嗦时，编码者的一致性信度就不高。如果某一类别包含很多组成部分或者包含有子类别，研究者最好对各组成部分都进行清晰明确的阐述。如果某一类别包含了很多组成部分，或者该类别所包含的组成部分在编码过程中出现频率很高，这样的类别应当继续分解为两个或更多的独立种类。编码者在编码过程中需要做出的个人判断越少，其信度就越高。

如果过于强调信度，可能会使得编码者过分地小心翼翼，甚至会深陷进数字的怪圈中，难以自拔，劳而无功。只有在确信了别人也会这样做成功之后，他们才会进行编码，至于样本中那些处于边缘地位的案例，他们就会忽略。如果因为态度上的保守而失去过多的信息，那么，对同一份文献进行重复编码是保证信度和数据丰富性的唯一方法。但是，如果这样做，就会增加研究的费用。研究者发现，对典型性样本进行重复编码是有益的，可以用“低信度”数据与“高信度”数据进行对比，以确定因为保守而造成的数据偏差的程度。

本章小结

内容分析这个术语，描述的是一系列研究方法，这些方法的使用有助于我们系统地收集资料、分析，并进行推理。因此，应当根据研究的问题来确定是否使用内容分析这一方法，或者从众多的研究方法中选择一个具体的研究方法。

研究结果，包括从收集到的证据中得出的推理结论等，是用来验证假设的。认真仔细地设计和实施研究，可以提高基于推理而得出结论的可信程度。

通过内容分析，可以获得各种各样的证据——包括质性证据、频数、趋势、意外事件和强度——我们可以从这些证据中推理出一些结论。近年来，内容分析研究也用于推导文本作者的个体特征。决策者的意图、注意的焦点，甚至心理状态也可以通过这种方法推理出来。使用内容分析方法可以推理出作者自己都没有意识到的东西，比如说，可以分析出决策者的挫折感和满意度。另外，还可以从个体通常的行为模式中辨别出其行为习惯的改变。

从同样的证据中，还可以推导出与作者的个性特征无关的东西。在政策情境主题中，可以通过一种使用频率不高的研究方法来重建决策者的思维链（end-means chain）。此时，单个决策者行为模式中的矛盾以及众多决策者之间的冲突，是很清楚的。

研究者，如果对根据单个国家或者数个国家的敌对程度的变化来进行推理很感兴趣的话，便可以预测出该敌对性的发展趋势，方法是将某一主题出现的频数或强度——要么使用其中的一个，要么两个同时使用——作为衡量指标。从敌对程度的上升或者下降中，分析家就可以对冲突发生的可能性进行推理。另一方面，研究者也可以使用意外事件分析的方法来确定两个变量指标之间的关系，如满意度和挫折感。从这些变量之间的关系中，研究者还可以进一步推论出关系的紧张程度。

第 4 章[①]

有关政策态度和行为强度的定量测度

测量中的基本问题

背景分析

在国际政治危机系统中，能够用来区分因子概念的有效方法之一是测量时使用 Q 检验的方法。在"个性评价和精神研究"中，为了促使主体的评价客观化，此时主要使用 Q 检验的方法。它最初出现于斯坦福大学进行的"国际冲突和整合（international conflict and integration）"项目研究中，作为有效而便捷的方法来测量描述危机的变量的强度。

在测量"1914 年欧洲危机（the 1914 crisis）"的参与者的语言表达中所包含的敌意的强度时，研究人员对于该方法的使用便超出了一般探究的范围。不久，它就成为非常有用的方法，既可以测量国际政治行为中所包含的情感因素，又可以测量其中包含的行动因素。人们认为，根据一系列可测变量，持续发生的所谓"紧张状况

---

① 本章节选译自罗波特·R·诺思（Robert C. North）等 1962 主编的公共政策研究论文集：《内容分析方法》之第 4 章。罗波特·R·诺思教授生前是美国斯坦福大学政治系的资深教授，他被美国政治学界赞誉为"开创国际政治文献定量分析方法的先驱者"。诺思教授已于 2002 年 8 月 2 日逝世，享年 80 岁。在他的最后岁月他宣布将他平生有关国际政治文献话语分析的论文论著的引用权和自己所有的版权均无偿公诸于世。M·乔治·赞尼诺维奇（M. George Zaninovich）当时（1962 年）担任斯坦福大学政治系诺思教授研究室的助理教授。本篇题目由本书编者在翻译后所加。本章由江雪梅、李钢翻译，李钢负责审校。

(tension profile)"可以通过Q检验的方法来重新构建,以进行比较和分析。其次,情感表达和行为表达的测量标准也允许人们对不同类型冲突情境中产生的偏离模式进行比较。再次,与国际危机相关联的假设和理论得到了发展,可以在一系列包含了冲突的历史事件研究中进行重新检验并限定其意义。

实践已经证明,使用一系列变量对表达的强度进行测量和比较时,Q检验方法是十分有用的;因此,作为一种有效的工具,Q检验可以帮助我们从简单频数分析转换到在强度水平上进行分析。因为有了这些强度水平的量值,人们便可以运用数学模型和代数公式来描述国际危机中的行为状态。

变量的维数

这种研究的初始目的在于确定危机所包含的可测变量或者测度一些特征的本质。变量的选择,很大程度上由特定研究的理论框架和假设所决定。

使用Q检验,保证(研究的)可测性的先决条件主要是变量的概念化,以及对变量进行处理,使之在限定的范围内按照强度从最小到最大的方式排列。我们不能认为其中点,或者说这样的维度(或连续体)模式是"中立的",就是说,不能认为它们表示的是零价值。与此相反,人们认为这种模式是呈现给评分者(评价研究人员)的、仍属于命题陈述(statement)领域内的、表现强度的一种普通方式。对于具有区分性的几个类别中的任何一类来说,使用Q检验并不要求具体的测量原则或标准,评分者只要能够根据已有的意义限定来判断特定变量之间的相对高或低就可以了。然而,由于缺乏任何形式的精确标准,评分者必须充分理解变量的规定意义,以便于他在所要测量的领域内,不仅能够对无数的变量进行测量,还要能够保持测量的前后一致性。

在Q检验中,对变量进行定义,并不能得到变量的维数,也就是不能确立该变量的正负两个端点;通过Q检验中的统计方法,即对任一变量进行从低到高的强度排列,可以确定变量的维数。因此,一个简单的变量,如敌意,可以从模式、平均值或特征的表达等方面来进行限定,而不能限定它的变化范围。对于可测变量来说,它的维数并不是定义的功能,而是对之进行测量的工具。

在某些测量方法中,主要用名词词组的形式来限定并描述变量的维度。例如,在"敌意—友谊"中,通过对两个有区别的概念——

敌意和友谊——进行限定，便可以测出这对变量的维数。该例中，根据经验，已知的这对概念在意义上是相互排斥的，这就使得在“敌意”和“友谊”之间建立经验型关系成为可能。然而，如果使用Q检验方法，按照Q检验中从低到高进行强度排列的规则，“敌意”这个变量便需要有自己独立的维度数目。假定用同样的方式处理“友谊”，在已知的例子中，我们可以发现“敌意”和“友谊”之间存在着一种积极关系，这个发现有利于我们进行特殊的分析。通过这种方法，研究者并没有失去有关信息，有关任意两个变量间的经验联系的信息。

简而言之，不能对变量进行狭义的描述，因为狭义定义不允许对变量进行处理，不能在内容上按照强度从低到高的形式对之进行排列。另一方面，变量的定义应当非常精确，以保证各个评分者的测量之间有较大的信度和较高的一致性。

基本测量单位

为了保证测量的精确度和信度，必须对基本可测的或记录的单位进行严格限定。研究者必须构建基本的测量单位，以保证其独立性，因为这些定义将影响到将要进行的一系列的测量评分工作。这一目标要求将复合的命题陈述简化为简单命题，就是在评分之前就把同质性问题归为一类。

对研究所要测量的基本的命题陈述单位（unite-statement）或主题进行限定时，应当考虑下面几个方面：①感知者（perceiver）；②被感知者（perceived）；③目标（target）；④描述性联结（descriptive-connective）。这四个因素决定了变量的本质[①]。正如我们在内容分析一章中所论述的，任何一个既定的命题陈述单位或主题都只能拥有上述四个因素中的一个。通过这种方式，研究所涉及的不同变量才可以清晰地区分开，每一个变量也才是独立的。

保证测量的透明度和精密度的一个基本条件是，任何一个命题陈述中只能包含一个描述性联结，否则的话，研究者将永远无法确定到底是哪一个描述性联结影响了评分。任何一个命题陈述中只

---

① 这里所说的描述性联结包括本书（指原书。—译者注）后面的章节中即将讨论的动作相关者（verb-connector）、共同意思评价者（common meaning evaluator）或者共同修改者（incorporated modifier）等因素。描述性联结显示了主题的三种基本类型：a. 动词形式，“X 攻击 Y”；b. 形容词形式，“攻击性强的 X”，这个主题也可以表述为“X 具有攻击性”；以及 c. 判断性形容词形式，“X 正在动员”。主题的这三种类型都必须作为独立的可测命题陈述单元来进行编码和处理。

包含一个描述性联结，保证了一个既定主题只提供一个变量①。如果我们让评分者来给一个相当复杂的命题陈述打分："狡猾的××人将军事力量运到××，目的在于消灭××人和对抗联合国军队。"对于"敌意"这个变量来说，研究者无法知道在"狡猾的××人"、"运送军事力量"和"消灭××人"三者中，主要影响评分的是哪一条。如果我们将命题陈述或主题进行编码并进行评分方面的准备，这个问题便迎刃而解了。上述这个复合命题可以分解成三个独立的小主题或命题陈述单元，如："××人是狡猾的"、"××人将军事力量运到××"，以及"××人要消灭××人"②。用这种方式处理复合命题，有助于在测量的命题领域中保持一致性，继而保证了测量过程的信度。

将一个复合命题划分成基本的主题单元，这一过程显示，描述性联结包含两个相互独立的种类，一是情感的（affective）或测量的（evaluative），二是行动（activity）或行为（behavior）方面的因素。相应地，任何命题陈述单元的内容都可以从感情和行动这两个基本方面来进行测量。在实践中，这两个基本类型的目标或目标表现形式大致相当，都是通过动词来表现的③。为了测量，每一个基本类别对于变量类型的要求是不同了，例如，敌意和满意是用来描述感情的，而说明（specificity）和暴力是用来描述行动的。不论是命题陈述单元，还是变量，其处理和选择的方式是一样的，就是避免引起这两种基本独立的类别的冲突。

同时，应该注意的是，对命题陈述单元样本进行处理时，不应涉及特定的国家、决策者和地理位置。对内容进行这样的"掩蔽"处理，为的是减少评分者可能带进评分过程的偏见因素。对命题陈述做掩蔽处理时，可以简单地用一个字母来代替某一国家，如用符号"X"来代表"奥地利"。例如，"（前）苏联已经动员起来反对德国"，可以转化为"X 已经动员起来反对 Y"。对每一个待测主题进行掩蔽处理，在一种白纸上打印出来，每一个主题的旁边都写上识别编号。还有一张是对主题进行编码时记录下的编码纸，记录有与各主题相关的其他信息，两张纸上相对应主题的编号是一致的。

---

① "是"的各种简单的或复杂的形式，如，"是"、"可以是"、"可能是"、"必须是"等，在实际测量过程中是有区别的，但在研究的编码或预备阶段并不专门区分这些词。

② 伴随着对消极因素的处理，也产生了一些问题，这些问题在实际测量前、在分析的准备或编码阶段必须得到解决。

③ 可以将副词结构的功能简化为加强主题的力量或强度，不专门对它们进行分类处理。

Q 检验中数值的分布

Q 检验测量方法的使用基础是已知领域中待测命题陈述以等级顺序排列。为了便利和经济的缘故，将这种等级排列的数据划分或“强制性”划分成九个各自不同的类别。因此，最终呈现在评分者面前的不是特定主题领域内的依等级排列的单个主题或命题陈述单元，而是依等级排列的命题组（rank-order of classes of statements）。这九种类别也可以看作是待测领域内命题陈述单元的分布状况。测量时的每一点在逻辑上都显示了一个类别数值（category-value），这个类别数值表现了双重功能。

命题陈述单元在九个类别上的分布，可以描述为，或者说是确保，评分者必须进行九种划分。Q 检验测量方法所使用的描述性数据分布如表 4.1 所示：

表 4.1　Q 检验测量方法描述性数据分布

目录(y)	1	2	3	4	5	6	7	8	9	(价值)
百分比(x)	5	8	12	16	18	16	12	8	5	(数量)

这种 Q 检验数据分布的主要特征就是对称或平衡，其图形表现为处于中间的种类是“凸起”的，数量最大。这种分布要求评分者在类别 1 和 9 上填入的数值是相等的，类别 2 和 8 上的数值是相等的，如此类推。

Q 检验数据分布的基础是一个假设，就是，在可以代表一个既定现象的样本中，大量的测量单位显示的是数据的“平均值”，或者平和而不是极端地描述了该现象的特征。这个假定认为，在 Q 检验中，数据的分布状态大致趋向为统计学上众所周知的“正态分布曲线”。然而，它也揭示了一些 Q 检验数据与统计学上正态分布明显不同的特征①。Q 检验中数据的强制性分布从任何意义上来讲都是不“现实的”或不“真”的，就是说，Q 检验数据的强制性分布并不存在着不容置疑的内在真实性。它只是一种简单的测量工具，用来督促评分者努力区分待测命题。

Q 检验数据的强制性对称分布使得变量之间的强度比较以及

① 例如，关于“通常的对称钟型分布”或正态曲线，从平均数正向偏离一个标准 1.00 到负向偏离一个标准 1.00 的范围内，涵盖了总数的 68%。然而，在 Q 检验中，同样的范围只涵盖了总数中很少的一部分。

变量在不同时期的强度比较成为可能。因为我们用类别值5来代表可测变量中的任何一个变量的表达形式,我们就有可能根据强度与5偏离的程度来比较各变量。例如,如果测量敌意和友谊的强度,在一个既定的时间段里,友谊的值是5.50,而敌意的值是4.30,那么我们就可以说在这种情境中,友谊的强度高而敌意的强度低;我们还可以认为人们对于友谊的描述要比对敌意的描述强烈。测量过程中,通过对不同强度数据进行强制性对称分布处理,把敌意和友谊放入一个共同的测量体系中,就达到了可比性。

另外,强制性分布也减少了评分过程中产生不适当的刚性标准和"绝对标准"①的可能性;它也可以避免测量过程中对评分者可能拥有的特殊分类倾向的歪曲。如果评分者将适用于某一命题陈述单元领域的绝对标准迁移到另一个领域中,这种迁移可能会误导第二个研究领域中命题陈述的相对等级排列。在Q检验中,绝对标准是根据某种统计方法来确定了,即,根据描述性分布来确定绝对标准,而不是通过建立刚性的标准将这些命题陈述排斥或包容在已知的类别范围。

需要测量的命题陈述单元或主题的范围大小随着研究目的的不同而有所变化。实践证明,命题陈述单元少到25,多到1 500,都可以成功地测量。待测命题的多寡涉及对选定的评分者进行控制的力量的大小,以及经过掩蔽处理的主题的清晰程度。为了确定测量表中到底填入多少条命题陈述单元,我们应当做个计算,即用既定领域中的命题陈述的总数乘以每一类别的分配百分比。

评分者应当记住,Q检验方法对这些命题陈述的区分是相对的,而不是绝对的,测量过程中,评分者必须将一条命题陈述与另一条命题陈述做关系对照,或者与其所面对的所有主题进行关系对照,而不是与来自于该主题领域之外的刚性标准进行比照。测量中的Q检验方法假定,在不同的命题陈述单元领域中,对于某一个既定变量来说,存在不同的表述方式,这是由测量过程决定的,而且只能经由许多评分者的一致认可才能确定。Q检验中数据的强制性对称分布既向这些评分者提出了必要的要求,也为比较不同变量的

① 术语"绝对标准"表示,确立了精确而固定的原则,将命题陈述排斥或包容在每一个不同的类别中。根据绝对系统理论,如果我们使用九分量表,整个研究领域,大概有100个命题陈述,那么我们可以令人信服地将这些命题填入九个类别的三个中。相应地,这种系统不是用来区分细微之处的,但是仍然可以测出待测命题之间的差别。Q检验方法解决了这个问题,方法是每一类别只允许填入固定数目的命题。

强度提供了一个共同的基础。

测量程序和信度

在正式施测以前,必须向评分者提供变量的清晰定义。这些评分者应当充分地对变量的定义进行讨论,做好思想上的准备。另外,应当从与研究有关联的命题陈述领域,而不是同一个命题陈述总体中选择几条命题来做一个测量的"预备实验(trial run)",这样做可以保证评分者使用同一标准对变量进行评分,而且为进一步判断评分者的信度和一致性提供了建议。在预备实验阶段,应当推导出初步的信度系数,对于那些评分明显不一致的主题,评分者应当重新审查并讨论。一旦正式测量开始了,评分者就不能彼此商量了。

为了便于操作,可以将测量过程划分为两个相互独立的阶段。在第一阶段,评分者应当在牢记分类原则的基础上,"自由地"给所有的命题陈述单元或者主题打分,然后填入九分量表中①。第二阶段,第一阶段完成后,便得到描述性分布数据的大致图形,此事,评分者应当重新阅读这些主题,修订他的初步测量结果。评分者重新阅读时,应当对初步测量结果进行必要的修正,目的在于得到进行Q检验所必需的强制性分布数据。在某些例子中,还有必要第三次、第四次阅读命题陈述。实践证明,仅仅在非常稀少的例子中,对于某一命题陈述的测量进行修订时,其在九分量表中位置上下移动超过一格。之所以如此限制是为了避免测量结果的过分改动,其假设前提是评分者最初对某一既定主题的测量是准确的。

一旦得到了合适的数据分布,根据其在九分量表中的位置,每一个命题陈述单元或主题便有了一个类别值(the value of the category)。接下来的任务,就是用恰当的测量数据形式,将这些命题陈述单元的类别值统计并记录下来。

数据的信度,由众多评分者对所有命题陈述进行测量所得的结果之间的一致性程度来决定,测量过程中起作用的三个因素决定了信度的高低。这三个因素是:①评分者所拥有的集中注意能力和注意力的跨度,②经过掩蔽处理的命题陈述单元的清晰性和简单程度,③对评分者进行的培训的有效性及持久性。通过计算评分者相

---

① 在把一个系统逐渐分解成三个部分的过程中,也可以使用其他方法。一旦最初的三部分确立了,接下来就是将每一个大部分分解成三个小部分,最终达到将系统分成九类。

关系数(inter-judge correlation)和合成信度(composite reliability)我们可以确定评分者的一致性程度。公式:

$$r = \sum d_{i,p}^{2}/K$$

可以用来计算评分者相关系数。其中,i 和 p 代表两个评分者。常数 K 由描述性数据分布和命题陈述单元的总数决定。在表 4.1 中,根据其描述的 Q 检验数据分布状态,计算这个常数的公式就是:

$$K = x_1(y_1 - y_9)^2 + x_2(y_2 - y_8)^2 + x_3(y_3 - y_7)^2 + x_4(y_4 - y_6)^2$$

在这个公式中,x 值指的是某一类别中所包含的命题陈述的总数,y 值代表了限定这一类别的顺序数①。例如,对于一个总数为 100 的命题陈述单元来说,其 K 值为 864,而总数如果是 500 的话,K 值就等于 4 320。K 的数值随着某一领域中命题陈述单元或主题的数量的变化,成比例地增加或减少。

评分者相关系数的可能变化构成了一个连续的数值范围,从 -1.00到 +1.00,只要相关系数值超过了 0.00,就可以认为这时的一致性程度要高于纯粹的随机测量。如果通过随机抽取的办法抽出数对评分者,计算出他们的相关系数,然后计算这些相关系数的平均数,那么,这个平均数的发展趋势是趋近于 0.00。一般认为,评分者相关系数为 +0.70,或者高于 +0.70 的话,评分者之间存在着非常高的一致性,也就可以用这些评分者的测量数据来进行分析。评分者相关系数显示,评分者对每一个命题陈述单元或主题进行测量时,由于相互之间的不一致而导致的测量值差别平均是 1.5 个类别(the average disagreement is in the order of 1.5 categories per unit-statement or assertion)。因为评分者必须在九分量表的九个种类中进行判断选择,这个系数表明了评分者之间的差别达到了相当大的程度。如果评分者之间没有达到某种一致性标准的话,就必须进行第三次测量,也许还要进行第四次测量,其目的就在于达成共识、强化数据。

这种测量的共识,可能发生在三个或更多的评分者之间,就是前面所说的合成信度。通过几个评分者中两两相关系数的平均数,

---

① 按照公式,可以这样计算:计算出第 1 类和第 9 类的差,第 2 类和第 8 类的差,如此进行;将差值平方;乘以每一种类中命题陈述的数量;最后计算出总和。

我们可以计算出这几个评分者的合成信度系数①。公式如下：

$$r = N \times 评分者相关系数的平均数 / \{1 + [(N-1) \times 评分者相关系数的平均数]\}$$

其中，N 代表的是达成共识的评分者的数目。合成信度的值总是高于单个的评分者相关系数。一般认为，合成信度系数达 +0.80，就足以保证分析数据的可靠程度②。假定有三个评分者，他们的评分者相关系数的平均数是 +0.65，这三个评分者的合成信度就是 +0.85。如果评分者在分析所有的数据时对每一个命题陈述使用的都是同样的测量方式的话，就可以计算他们的合成信度系数。

测量过程中得到的类别值，来源于测量的命题陈述单元。可以将这个数值填入一张基本数据表格或者 IBM 卡中，在这些表格中，记录有与涉及既定主题的其他相关信息资料。为了便于分析，针对某一特定的命题陈述，可以使用不同的测量方法。

分析测量数据

对数据进行分析的过程中，产生了两个问题：①我们选择哪些类别值来进行分析，②我们将使用什么样的标准。研究者必须根据特定的理论和相应的假设做出判断。

我们可以用一个简单的公式来确定已测变量的强度程度。如果 V 代表的是变量的表达程度或强度，那么：

$$V_{ij} = \sum (d\,c)/N$$

其中，i 代表涉及的因素（例如，一个国家或者一个特定的决策者），而 j 代表了确定标准而持续的时间段③。dc 的值是对既定的命题陈

① 此处所说的信度，意味着将该合意（或平均）分数与另一个由相等的评价者作出的合意（或平均）分数相对照时所得到的一致性。就是说，如果我们从同样的评分者总体中再抽取出一个评分者样本，得出第二个评分者一致同意的测量值，将这第二个合意数值与从第一组评分者方面获得的合意值进行相关计算，计算结果就是我们这儿所说的信度系数。

② 将这种合成信度作为标准，首先是由布洛克博士提出的。“同时，因为我们选择的是平均分数，根据简单的心理测量学的逻辑，我们可以得出十分令人信服的结论，就是：在评分者一致同意的测量中，有效性的累积速度大大超出了错误的累积速度。总地看来，观察者的个人气质、疏忽大意以及其他缺点是相克的，最终获得绝对真理。对合意有较高效度的希望可以得到实际经验的支持，这些经验来自于研究中调查的所有例子。”

③ 因为 Q 检验测量中数据的分布是对称的，很明显，从“时间 1”开始到“时间 n”为止，就是说，对于所有待测命题陈述单元来说，分析中涉及的所有国家的 V 值必须规定为 5.00。在这样的情境中，公式中的 N 值将等于待测的所有命题陈述的总数。

述单元或主题进行描述性联结处理时所赋予的类别值。因此,这个符号代表了变量强度的单一测量值,而 V 代表的是经过公式计算的累积值。还必须对包含在 N 中的命题陈述单元进行处理,处理方式包括:与已有的成分——感知者、被感知者和目标——进行联结,使用其他基本信息,如原始命题陈述的数据、沟通工具的类型等,对之进行限定。例如,假定一个人想确定 X 国家从"时间 1"到"时间 5"这段时间内以"敌意"作为目标的分析标准,那么就可以使用下面这个简单的代数公式:

$$H_{x1} = \sum (dc)_{x1}/N_{x1}, H_{x2} = \sum (dc)_{x2}/N_{x2}, \cdots$$

其中,下标 1,2,…,5,代表的是时间段的顺序①。所得的值 H 以及相应的 N 的值,最大限度地显示出这个时间跨度上的波动值。分析的结果可以为我们提供一系列指标,表现了"敌意"这个变量的强度趋势。根据研究者获得的某些特殊信息,也可以容易地推导出其他公式。

批评意见指出,只有当其测量的命题陈述属于同一领域时,才可以使用 Q 检验方法。某种意义上说,这个结论是正确的,就是说,如果需要进行强度比较的命题分属于两个独立测量的领域的话,因为通过最初的定义,这两个命题形式上的数值和平均数都将是 5.00 个类别值。那么,在不同时间里对两个领域进行分析,就需要建立一个新的、共同的概念领域,以便主题与主题之间有可比性。方法可以是,从两个领域中各自抽取类别值为 1,5 和 9 的主题组成新的样本,对这个新样本里的所有主题进行重新测量②。这种重新测量完成后,从两个原始领域中抽取的命题陈述就可以保持独立,同时,通过计算两个样本的平均数还可以得到全新的评价指标。

从另外意义上说,这种批评又是无效的。如果表示连续时间内强度变化的曲线的运动是相互关联的,那么经过独立测量的两个主题领域内的变量就是可比的。在本例中进行比较的不是强度的形式上数值或者一般水平,而是代表了强度变化的曲线的运动方式。进一步说,在对一系列变量的内在关系模型(即矩阵)进行比较时,

---

① 按照公式,可以这样计算:找出 X 国的所有包含变量"敌意"的命题陈述,作为时间 1 上的目标,计算出每一类命题陈述的类别值,除以研究中涉及主题的总数。结果就是一系列指标,根据这些指标可以画出一个曲线,代表了 X 国的"敌意"强度的变化状况。

② 这种对独立测量的两个命题陈述单元或主题领域进行比较的方法是布洛克博士在 1962 年 5 月 12 日召开的一次研讨会上向斯坦福"国际冲突和整合"研究小组提出的。

这样的批评是站不住脚的。这些模型并不决定于这些命题在待比较领域内的形式上的强度水平，而是由这些主题的曲线从一个时间点到另一个时间点的运动方向和它们在某些时间段里可比的强度水平所决定的。在相关分析和模型分析中，不必建立新的统一的领域，就可以对两个独立的相互冲突的情境进行比较分析。

对于某个特定的分析来说，可以针对“判断活动”和表达的感情强度之间的关系来确立特殊的假设。使用Q检验测量方法，对有关联的行为变量（明确表达和暴力）和情感变量（敌意和挫折）进行独立的测量，可以得到二者在某一共同时间段上强度趋势的相关。例如，通过简单的相关分析和偶然性分析，既可以验证，也可以推翻“在冲突类型A中，作为‘调整活动’的暴力发生在高强度的挫折感之后”这一假设①。与此类似，也可以确定某一次国际危机中消极情感（敌意和挫折）与积极情感（友谊和满意）之间的强度比率，这为进一步确定奖励和惩罚的平衡奠定了基础。这些例子仅仅提供了一些建议，帮助我们理解怎样使用测量技巧来确定并验证假设和主张的稳定范围。

Q检验可以帮助研究者对一组变量在一段时间内的强度水平进行比较和联系。我们可以通过两者不同的方式来使用比较分析：①对处于同一冲突情境中的不同变量进行比较，以及②通过确立不同变量之间的关系模型，对处于冲突情境中的变量进行比较。这种确定不同变量的强度或表达程度的方法，可以用来验证假设和观点，这些假设和观点是对国际危机中各国行为习惯的描述。

总　结

Q检验测量方法，通过计算许多聪明而感觉敏锐的评分者之间所谓的“主体间一致性”，尝试着确立供进一步分析所用的数据的信度。为了应用这种方法，最初是在统计学意义上设定的假设，即，假定测量的命题陈述单元的数据呈描述性对称分布。对量表中任一种类的内容进行测量时，就不再设定任何假设了。这样做的目的是为了确定Q检验量表使用时的相关性质，而这是不需要对每一种类进行硬性规定的。

---

① 在这里使用“冲突类型”这一术语的目的在于表明，根据行动变量和感情变量之间的特殊关系，我们可以确立冲突情境中的类型学。广义地讲，我们可以定义冲突的两个基本类型，积极感情—消极行动，以及消极感情—积极行动。

Q 检验测量方法的最大优点就是它经济简单,便于使用。实际中,只要对评分者进行短期培训,便可以在相对短的时间内完成 Q 检验。另外,Q 检验强调主题所处的情境和该主题的主旋律,因此可以说,对于命题陈述的测量是全面的。它还强调所测命题陈述在其领域中的独特关系。因此,在辨别细微的差别时,Q 检验尤其有效,而其他精心设计的方法可能会不经意地忽略了这些细微的区别。最后,通过 Q 检验,我们可以对变量进行有意义的分析和比较,因为强制性数据分布为处理不同变量提供了一个共同的基础。

Q 检验测量方法为我们提供了确定描述危机的各种变量的强度,并将之转化为数量形式的工具。进而,它可以用来测量世界政治领域内的情感变量,如敌意和挫折感。将来,我们可以使用 Q 检验来测量和分析国家间危机的其他方面。

Q 检验的使用

深入思考 Q 检验

本章前半部分,我们介绍了一种测量技术,这种方法在心理学研究中曾经取得了一定的成绩,在对国际关系中的问题进行比较研究时,这种方法也是有效可行的。随之而来的问题是:Q 检验方法适用于国际关系中的哪一类或哪几类问题?要回答这个问题,仍然需要讨论 Q 检验运用的长处和短处。

我们的目的不是叙述 Q 检验的发展历史,不是叙述 Q 检验产生的文化背景,在那个文化背景中,术语缺乏一致性,人们经常进行一些晦涩难懂的技术性争论。然而,如果对 Q 检验的产生发展历程进行简要的梳理,则可以帮助读者理解 Q 检验的可能应用范围。从根本上说,Q 检验修改了常用的相关设计方法。当一组被试在两次或多次测验中的成绩相关时,我们常常使用这种相关设计方法,此时,这种研究设计包括个体间的相关测试。对这种设计方法加以修改——继而产生了 Q 检验方法——可以计算两个个体在一组测试中的成绩之间的相关。后一种情况下,我们就有可能对两个特定个体的行为进行比较。这种修改也引出新的问题。在最初的研究设计中,因为在任何一个测验中,个体与个体之间的成绩分数是可比的,所以我们可以将一组个体在两次测验中的成绩相对照,就是说在同一次数学测验中,可以将 10 分与 50 分进行比较。在修改过

的研究设计中，这种比较就无效了。因为每次测验都有不同的测量单位，测验的分数不再具有可比性。为了解决这个问题，才引入了Q检验。如果Q检验中的每一个指标都是针对某一次测验的，然后让某一个人来进行Q检验，那么当那个人进行一系列测量时，这些指标就是可比的了。明确地说，对Q检验的每一个指标进行强制性分布处理，那么这些指标的最终得分是可比的。因此，数据呈强制性分布的Q检验的产生，保证了不同测验成绩的可比性，使得计算个体在不同测验中的成绩的意义相关成为可能①。

Q检验的使用

对于Q检验数据的使用方法，人们做了很多讨论，其中最好的论述是杰克·布洛克(Jack Block)做出的。他提出的第一个用处就是，一组指标可以经过两个或更多个体的Q检验处理，那么结果就是可比的了。通过这种方式，可以对几个主体的反应进行比较。例如，对于同一个主体来说，可以将他真实的行为与他理想中的行为进行比较。两个医生可以为一个病人看病，或者两个心理学家关于某个概念(如创造性)的指导性意见也是可比的。

Q检验数据的第二种用法，是比较两组个体的Q检验值。例如，我们希望把低收入黑人组的Q检验值与低收入白人组的Q检验值做比较，以发现这两组在各项指标上的显著差别。

第三种使用Q检验数据的方法是，将来自于一组或多组的Q检验值与理想的Q检验标准进行比较。我们可以先让心理学家为精神分裂症病人确定一个理想的Q检验值，然后用一组精神分裂症病人实际Q检验值与之作对比。或者，像布洛克所说的，"我们可能希望了解，与其他组相比，某一组个体的Q检验值是否更接近标准。"

最后，布洛克指出，还可以用Q检验数据来确定或限制被试的类型或组别。"以前，我们都是根据一些独立的标准，将被试分组，然后分析各组被试的Q检验值；现在，我们可以将顺序颠倒过来，我们可以根据被试的Q检验值分组，然后分析信息的独立来源以计算组员之间的相关。"这种方式可以确定数组主体所共有的基本

---

① 然而，有人指出如果使用了检验程序，就没有必要对数据进行强制性分布处理。如果一个个体不考虑数据的分布状况使用同一个量表对Q检验指标进行分类的话，就有理由相信，各指标的分数之间存在着意义相关，因为它们是由同一个体按照同样的维度来分类的。

特征。

也许,将Q检验应用到国际关系中最明显、最简单的方法涉及一些包含了心理学因素的研究问题。例如,我们要比较一组和平主义者的Q检验值与另一组军国主义者的Q检验值。Q检验的指标可以选择个性特征或美国外交政策的表达方式。这种设计基本上属于布洛克提出的第二种使用方法,不同之处在于我们研究的是国际关系变量,而不是心理学变量。

与此相类似,Q检验还可以与游戏相结合进行模拟研究。我们可以同时模拟创造两个情境,一个包含国际危机,一个不包含国际危机。然后要求参与者对一系列指标进行Q检验测量,如:"X不公正","X不懂外交"或者"X说谎",接下来比较不同参与者在两种情境中的测量。

进一步的修改与布洛克的第一种提法相应。可以将Q检验运用到国际关系中来澄清某个关键概念。我们可以构建一系列Q检验指标,包括:"X断绝了与Y的外交关系","X对Y实施贸易壁垒",以及"X的新闻报道谴责Y的入侵"。然后请两个评分者对这些指标进行测量,测量这些命题陈述中所表达出的敌意的程度。如果评分者是随机抽取的,那么,比较其结果可以看出外行人怎样使用"敌意"这个概念。如果我们从拥有国际法知识背景的人中选择评分者,那就有理由相信该结果表示了两个权威之间的一致程度或者不同程度。进一步说,如果这两个评分者是从拥有国际法知识背景的人中随机抽取的,我们甚至可以认为结果的一致性程度代表了国际法领域内的总体一致性。

Q检验的优点和缺点

如果对某一个情境中的变量按照规定的维度进行测量的话,Q检验是否就一定适用呢?换句话说,Q检验最适用于哪些类型的国际关系研究设计?对这些问题的回答,需要进一步讨论Q检验的优点和缺点,需要考虑Q检验的不同使用方法。

很明显,如果根据强制性准正态分布的假定而无法得到研究问题所必须的资料时,Q检验就是不适用的。例如,假定我们用几个变量——如敌意、友谊、挫折和满意等——的强度分布状况来描述一种危机情境。如果根据Q检验理论来测量每一个变量的强度,这种分析就是不可能的,因为,每一个变量的分布状态已经预先设定为正态分布了。在这些例子中,就应当使用另外的测量工具。

现在我们来讨论Q检验的长处和短处。

1. 一旦对一系列命题陈述进行Q检验的工作完成以后，每一命题的测量值便反映了其在已知的命题领域范围内的相对强度。由此，如果扩大或缩小命题陈述领域的范围，或者干脆从中选取一部分作为新的命题陈述领域，那么人们对同一命题的测量结果是不同的。应当注意的是，同一命题的不同测量值只有属于其测量领域时，才有意义。例如，假定有两场国际危机，我们有兴趣对这两场国际危机中存在的"敌意"进行强度的比较，从危机爆发开始，到危机达到高潮为止。如果我们独立地测量每一场危机中的敌意的强度，继而进行比较的话，就大错特错了。如果我们期望对这两场危机进行的比较分析富有意义的话，我们必须将两场危机中的所有关于敌意的命题陈述混合起来，一起测量。

确定相对值的测量技术对于很多研究问题来说，都是有用的。相对测量的唯一缺陷就是，我们不能用"绝对化"的术语做出概括性结论。例如，我们不能说，上述两场（或更多）危机中，敌意的强度从4开始，顶峰值是9，然后逐渐减少到7。另一方面，如果我们有足够大的危机样本，其中的每一场危机中都有一个行为习惯与我们描述的相似，我们就可以得出结论：危机所包含的敌意的强度随着时间呈曲线变化趋势。

尽管Q检验确定的相对数值对于多种问题的研究都是有益的，但这并不是Q检验所特有的长处。例如，如果让评分者按照强度从小到大的顺序，将50个命题陈述分为九类，并不对每一类中命题陈述做数量上的限定，那么所得的结果就是一个"相对"量表。进一步说，扩大或压缩命题陈述的领域范围的能力使得人们可以转而使用其他测量工具。因此，如果仅仅是想得到相对数值的话，其他的测量工具也是可以达到目的的。

2. 如果对每一种类中命题陈述的数量不加任何限制，就要求评分者进行测量并分类的话，也会产生偏见。例如，不同的评分者，敏感性会有所不同。一个评分者可能只对极端敌意的命题陈述敏感，而另一个可能对中等程度的敌意敏感。由此，两个评分者对每一个有关敌意的命题陈述的测量值就会存在着很大的不同，降低了测量过程的信度。因为我们是要对敌意的强度进行总结并得出一般性结论，而不是要描述特定的评分者的敏感性，我们就需要一种能够最大限度地减小这种个体偏见的测量工具。通过规定每一类别中命题陈述的数量，我们可以有效地保证每一个评分者的敏感性保持

在同样的水平上,而且以同样的方式显现出来。这种客观化过程实际上是Q检验所使用的准正态分布的一项功能。不论这种强制性分布是近似于正态曲线,还是双峰曲线或偏态曲线,我们都可以得出同样的结论。

假定因为数据或者特定问题的原因,我们不能使用任何形式的强制性分布,还有没有其他的办法可以保证数据的客观化?这就需要有一种测量仪器,即工具,类似于尺子,帮助人们进行可靠的测量。这样有用的方法当然是有的,其中最简单的就是,从待测的命题陈述领域中选取不同命题,对之进行测量分类,使之在量表中处于不同类别,作为样本提供给评分者。

3. Q检验中对数据做出的强制性准正态分布的假设,促成了等距量表的产生。这证明我们选取的某些命题陈述是有存在根据的,我们就是对这些命题陈述进行Q检验测量而获得的研究数据。另一方面,也可以做其他的假设,这同样也可以制成等距量表或其他量表,如比例量表、绝对量表。如果根据数据制出的是等距量表,而研究者希望涵盖整个Q检验过程,那么就可以使用其他方法进行测量,然后使用恰当的统计方法将之转化为正态分布。

使用Q检验的困难之一是时间的长度问题和工作量大小的问题。对于Q检验测量的进行来说,一张纸上最好只写一条命题陈述,而且要对每一条命题陈述进行掩蔽处理。因为限定了每一类别中填入的命题陈述的固定数目,所以一旦某一类别的命题陈述达到了规定的数量,那就必须将之移到另一类别。如果命题陈述之间彼此独立的话,这样做是很简单的。如果待测的命题陈述很少,那么不需要一张纸上只写一条命题陈述,也可以完成"移行"工作。但是,大多数例子中命题陈述的数量都很大,每一条命题陈述又必须进行掩蔽处理,对于"原始数据",我们就不能拿来就用。如果有数千个命题陈述的话,单是Q检验过程就须消耗相当大的资源。

4. Q检验过程的某些方面会降低测量的信度。Q检验要求评分者将N条命题陈述中分为九类,而且每一类别都有数量上的规定。面对大量的命题陈述时,评分者的辨别能力趋于降低,因为,正如布洛克所说的,在最后的分析中,评分者是对所有可能的命题陈述进行配对比较,当一个人面对越来越多的命题陈述组时,测量过程的信度就会降低,因此而限定了可以测量的指标的数量。

另外,使用强制性分布也可能导致测量过程发生其他错误。思考在一个极端情境中,划入某一类别的命题陈述已达到规定数额。

如果这时评分者发现有一条命题陈述与该类中相近,他该怎样处理这条命题陈述？这条命题肯定得归入其他类别,要么高一格,要么低一格,但是不论是把它划入高一格还是划入低一格,都没有逻辑依据。在这个例子中,两个评分者对这个问题进行相同处理的几率是很低的。很明显,这种情况发生的次数越多,评分者的信度就越低。

总而言之,最初设计Q检验的初衷是为了解决心理学研究中的特殊问题。因为假定它不受变量分布形状的影响,它便可以应用到其他的研究设计中。当研究设计不需要对数据进行强制性的准正态分布的限制处理(虽然这样做并不歪曲数据的真实性)时,是否使用Q检验,必须取决于研究者对Q检验的优点和缺点的衡量。仅以此作为告诫,告诫研究者仔细选择研究方法,使之适用于研究设计。



第Ⅱ部分

内容分析方法在公共政策研究中的应用

在第Ⅱ部分中，我们要将第Ⅰ部分中所介绍的内容分析方法付诸于尽可能各色各样的案例实证分析。所选取的5篇文章其分析的主题内容主要涉及中美国际政治关系、美国总统竞选报道、中国教育政策执行过程中的群众信访，以及美国267所学校办学政策的理念层面的分析，还有中国进入21世纪以来社会公众话语构成的当代社会舆情中，有关中国国家教育政策的社会话语内容分析。本书之所以选取了这些诸多主题并不太集中的政策内容分析案例，旨在向读者展示内容分析法这种起源于20世纪的政策方法，由于其在有关文献计量方法上的可靠性和编码的简易操作性等特征，对于话语文献内容分析的方法建设，在21世纪仍然具有广泛的发展运用前景；这意味着内容分析方法适用的范围比较宽泛，并且随着计算机运算速度越来越高效、信息技术的飞速发展，内容分析方法将会在公共政策执行过程、政策评估和政策科学决策影响变量的定量精确测度中发挥出越来越有影响力的作用。读者在阅读本部分所收编的5篇案例分析报告中，需要注意其中的这样几个方面：首先是内容分析方法的适用性。即哪些材料是适用于内容分析方法的，以及各篇分析报告的研究者是怎样依据不同的分析主题要求，选取内容分析方法进行不同切入点的内容分析的。再者就是内容分析方法的程序体现。文中的若干表格和分析步骤是如何按照主题内在的语义逻辑相互配合，来一步一步地实现对于分析主题的内在本质语义的层层揭露。接着就要注意研究不同的语料对于研究者确定政策分析目标的影响。为此本文选取了有关政策过程的动态信息和静态信息的不同语料类型（例如新闻、报纸、政府公报、学校的校训和群众向政府部门投诉的

"公众来信",以及政府公开的政协提案等不同社会阶层有关政策的话语文本)进行内容分析,以表示内容分析在探索公共政策全过程中具有非常重要的实证检验意义。内容分析是一种真正的公共政策分析方法,反之亦然,经得起内容分析的公共政策才能算是一个比较成熟、正确、有所作为的公共政策。

术语解释

[**命题陈述**(statement)] 即对于要研究问题客观、概括的提炼和表述,将所要分析的问题按照 3W(what, why, when)1H(how happen)等的政务格式表述清楚。

[**信度**(reliability)] 信度是指使用相同研究技术重复测量同一个对象时得到相同研究结果的可能性。信度并不能够代替精确度,信度不能够完全保证准确性,它指的是测量方法的质量,即对同一现象进行重复观察是否可以得到相同的资料。在对于政策实施内容分析方法中遇到的信度问题主要是分析人员的信度,即如果分析人员缺乏起码的内容分析技术训练,而又盲目自信,先入为主地对于大量有用的分析语料草率地进行分类,并且认识不到其所做的分析是对于政策真实性问题调研的破坏,并由此得出一些主观成分多的政策建议,就会为政策过程的进一步评价埋下隐患。

[**效度**(validity)] 效度是指实证测量在多大程度上反映了概念的真实含义。比如,最有效的测量可能永远也找不到,但是可以依靠表面效度、标准效度、内容效度和建构效度来综合分析效度。建构效度的基础就是指建构变量之间的有效逻辑关系。

[**语料**(analysis materal)] 指被选定分析的话语文本和材料。

[**政策评估研究**(evaluation research on the policy)] 政策评估研究的重点是专注于分析政策问题,拟订并监控政策研究设计、确定研究框架并评价研究政策可能的效果。将政策的全过程视为一个项目的完成过程。评估研究的目的是确定整个项目效能的过程。评估研究要求对项目或项目中的特定问题进行仔细探求,试图查明某一项目或组织中执行的策略——针对单一目标或多目标进行实现可能性评价研究。发现项目的价值、质量,并对是否继续进行项目做出建议和对于有问题的项目进行必要的改变。

第5章[1]
国际关系中政策热点的“对偶比较”测量

前文中我们论证了一个问题，即对于某些研究问题来说，Q检验并不是一种适用的测量方法。假定现在有一个问题，不能进行Q检验，而我们必须对它的变量进行测量，我们可以使用什么样的方法呢？此时，我们有两种方法可以选择，一种是第四章中介绍的方法，另外一种就是本章要讨论的主题。

要求对国际关系中的若干变量进行测量

为了更好地进行描述，我们将这种测量方法称为“对偶比较测量”。严格地说，“对偶比较测量”这样的说法并不准确，可能会引起歧义。我们简单地使用“对偶比较测量”这个词，目的只是想方便于人们来准确地记忆，便于人们记住这种方法。对“对偶比较测量”进行细致的研究之后，就会发现，这种测量方法并不意味着对测量的所有命题陈述(statement)均可以进行配对比较研究，而是从大量测量数据中随机选择一个小样本，仅仅对这个小样本中的数据进行配对比较。在具体说明这个方法之前，让我们看看“对偶比较测量”存在的理论前提。

国际关系中存在的大多数研究问题，都涉及非人格化的变量，即，这些变量并不能说明人与人之间的差别。因此就要求，在测量

[1] 本章选自迪纳·A·津尼斯(Dina A. Zinnes)等人在20世纪60年代的一些研究论文。迪纳·A·津尼斯曾与罗伯特·C·诺思(Robert C. North)有过密切的科研合作。后来(大约是在20世纪70年代)津尼斯到北伊利诺大学计算机工程系任教。

方法的使用过程中,人们可以对这些相关指标进行测量的准确程度进行判断。此时,对测量信度的要求很高。在这里,我们不打算叙述评分者之间的个性差异,我们关注的是他们对命题陈述进行判断的方式之间的差别,前提是他们一致同意使用某种测量方法。如果评分者是随机选择的,我们可以假定前一种差异的数量很小,可以忽略不计。另外,不论测量进行到什么时候,我们都可以对这个假定的效度进行测试。这就是众所周知的检验该测量的信度。我们选择两个或两个以上的评分者,让他们对一系列指标进行测量,然后比较他们所得的结果;通过恰当的公式,我们就可以得出他们之间相差的数据。如果这些差别属于可接受的数据范围,我们就认为评分者们能够客观地做出了判断,测量也是可信的。

如果评分者没有得出可信的结果的话,就麻烦了。继而产生的问题是:信度不高的原因是什么?导致低信度产生的原因至少包括以下两个方面:

第一个原因是,个体与个体之间存在差异。第二个原因是,在对这些指标进行测量时,个体缺乏判断的前后一致性。前面我们已经讨论过第一个原因了,也许我们应当进一步讨论第二个原因。例如,如果评分者在测量过程中,对他所要判断的变量没有清晰的概念认识,就会与其他变量相混淆。此时就产生了上述第二种状况。这种情况下,评分者是按照单一维度测量一个多向度的概念。很明显,对于该指标的测量结果就缺乏了前后一致性。某一指标的测量值是3,另一指标的测量值也是3,这两个3并不能直接比较,因为测量时,这两个指标分别处于不同的维度下。所以,测量过程中,有必要使用一些技巧,以消除个体偏见,并且暴露评分者测量的不一致性。

现在,我们来看测量过程,我们来看前一章叙述的Q检验中使用的测量过程。该测量中,要求评分者将一系列指标填入一张包括九个类别的表格中,结果这一系列指标在九个类别上呈现出"准正态分布"样式。因为反对Q检验的主要理由是因为我们将数据的分布状况进行了这种强制性处理,那么,抛开这些强制性分布的限制,我们继续进行测量。现在,评分者评价指标的强度,将之归入某一类别,此过程中,并不对纳入某一类别的指标进行数量限制。由此我们得到一份新的测量表,这种表适用于那些不能进行Q检验的问题。但是,有所得必定有所失。因为我们想起了另外一个问题:在数据呈现强制性正态分布(或者任何形式的强制性分布状

况)时,评分者测量过程中表现出来的偏见是一致的。因此,新的测量技术需要数据呈现另外的分布状态。

然而,需要强调的是,Q 检验并不强求将评分者评价的先后不一致暴露出来。对两种不同情况下同一名评分者对同一系列指标测量的结果进行信度系数的计算,并不必然导致这个结果。评分者在两种情况下的测量可能是不一致的。另一方面,评分者在两种情况下进行的测量可能是一致的,可是与其他的评分者相比,他的信度仍然较低。因此,信度系数并不能充分地说明测量的不一致性。

我们提倡的方法是对偶比较判断。这种方法与简单的测量程序——对指标进行等级划分,然后归入九个类别中———起使用,可以解决我们在讨论 Q 检验过程中产生的困惑。这种方法可以在不牺牲信度的前提下,运用于不适宜使用 Q 检验的问题。另外,它还可以暴露出评分者测量中的前后不一致。

对偶比较测量

也许,通过一个例证,我们可以更好地理解"对偶比较测量"。假定,在某一个领域内,存在 2 000 个关于敌意的命题陈述需要我们对之进行测量,以确定其在表达中的强度。我们从中随机选出 18 个命题陈述,写在纸上,为每一命题陈述编号,以便识别。接下来,在另一张纸上,我们列出这些命题陈述所有可能配对。然而,这个名单不包括这些命题陈述与自己的配对,比如,不包括 1—1 或 2—2;同样的配对也不能出现两次,尽管两个命题陈述出现的次序不同,比如,1—2 和 2—1 不代表两组不同的配对,它们实际上是一组。如果我们只取四个命题陈述,名单上的配对形式如下:

1——2
1——3
1——4
2——3
2——4
3——4

总的配对数由命题陈述的数量决定,计算公式为:

$$n(n-1)/2$$

因此,如果有 18 个命题陈述,那么总的配对数为:

$$18 \times (18 - 1)/2 = 153(对)$$

对偶比较研究包括下列步骤。我们认为每一配对都是独立的,然后提问:(我们的例子)每组配对中,哪一个命题陈述代表了表达所显示的敌意的最大强度?一旦确定,“赢家”,即最富有敌意的命题陈述,得到1票。在列出所有配对的纸上,添上“赢家”一栏,将这1票记录在案。对每一组配对重复这一程序。对上述4个命题陈述进行如是处理,我们可以得出:

主 题

配 对	“赢家”
1—— 2	1
1—— 3	1
1—— 4	4
2—— 3	2
2—— 4	4
3—— 4	4

接着,对每一命题陈述的票数进行统计,按照命题陈述的顺序号进行排列。上述4个命题陈述的排列顺序是:

命题陈述	票数(votes)
1	2
2	1
3	0
4	3

这里,我们回过头去看看“对偶比较研究”期待解决的困难。看第二个问题:“怎样暴露出评分者测量的前后不一致?”如果计算出每一命题陈述所得到的票数,就可能确定测量的一致性。如果是按照一个纬度对变量进行测量的话,即如果评分者使用同样的定义对变量进行限制、前后一致地对该变量进行测量的话,这些命题陈述的得票数将呈现出等级顺序。换句话说,我们能够根据这些“票数”对命题陈述进行分等,命题陈述与命题陈述之间较少纠缠不清。在我们的例子中,一共选了4个命题陈述。

可以看出，评分者在测量时表现出了完美的一致性，任何一对命题陈述都区分得很好，我们可以按照每一命题陈述的得票数，按照从多到少的顺序，排列出这4个命题陈述的等级来。

命题陈述	票数的次序(number of notes)
4	3
1	2
2	1
3	0

如果样本数达到18个命题陈述，可以认为，即便有2～3个命题陈述区别得不是很好，这个评分者的前后一致性仍然是非常高的。

我们认为，如果概念纠缠、区分不清的命题陈述的数量太多，评分者就没有按照既定的维度进行测量，或者是他的测量的前后不一致。然而，只有当我们最初就假定需要测量的命题陈述是可区分的时候，这个结论才可信。如果与测量变量相关的一系列指标是不可区分的，由于评分者没有办法确定两个指标中的哪一个强度更大一些，就可能，甚至必然导致命题陈述的纠缠。如果无法区分的命题陈述的数量属于可接受范围，也可以认为是体现了指标间的可区分性。

如果发现两个评分者信度都不高，研究者面临的问题就是，必须判断其产生的原因或者辨别出概念不清的指标的数量。基于先前的讨论，我们建议使用"对偶比较研究"的方法，增加指标间的区别程度，进行多维度测量，以确定问题产生的根本原因。这两个具体步骤可以通过对指标进行检查来完成。

现在假定，在我们所取的包含18个命题陈述的样本中，评分者的测量是前后一致的。继而产生的问题就是，使用"对偶比较研究"是否可以帮助我们解决评分者与评分者之间的信度问题。前面我们曾经提到，在Q检验中，为了解决这个问题，我们对数据的分布进行了规定，即规定数据呈现出强制性分布状态。在使用"对偶比较研究"这一测量方法时，我们可以使用一致性测验的结果。因为，对18个命题陈述进行测量，测量其表现出的敌意的强度时，所提取的数据呈现出等级排列(或近似于等级排列)的顺序。因为

这 18 个命题陈述是从 2 000 个命题陈述中随机抽取的,如果我们使用十八分量表,将敌意的强度分为 18 级的话,那么这 18 个命题陈述就分别占据了这 18 个等级。如果我们使用的是九分量表,将敌意的强度分为 9 级,那就有必要将上述 18 个等级合并成 9 个。这很容易,只要将 18 级中的每两个等级合并成一个大的等级,即原先十八分量表中的一级和二级变成现在九点分表中的一级,如是而已。

合并过程完成之后,九点量表中每一等级就有两个命题陈述样本。用这个结果作为"尺子"或"定义",我们可以测量剩下的 1 982 个命题陈述。步骤如下:从 1 982 个命题陈述中选取一个命题陈述,比如"X",与"尺子"进行比较:命题陈述 X 与哪一对命题陈述最相像? 对这个问题的回答决定了命题陈述 X 的测量值。

在对所有的 1 982 个命题陈述进行如此的比较之前,有必要计算评分者的信度系数。方法是,从这些命题陈述(2 000 个)中再抽取一个样本,这个样本不应该包括一致性检验中已经使用过的命题陈述,样本容量大约为 30 ~ 50,让两个评分者仿照前例对这个样本进行测量。参照任何一本初级统计教材上常见的信度计算公式,都可以计算出这两个评分者之间的信度系数。如果只有一个评分者测量了所有的命题陈述,那么只要证明他的测量是前后一致的就可以了。因此,本例中,必须计算同一评分者不同时间对同一组命题陈述进行的两次测量之间的信度系数。如果我们的研究中有不止一个评分者,还必须计算评分者之间的信度系数。只有信度系数达到要求的水平,该测量程序才可以扩展到所有的 2 000 个命题陈述(十八分量表被合并成九分量表后,原先的 18 个命题陈述的测量值应当相应发生变化)。

将十八分量表合并成九分量表,在九分量表中,每一等级上就有两个命题陈述,也有可能每个等级上命题陈述超过两个。本例中,从剩下的 1 982 个命题陈述中再抽取一个样本容量为 18 的样本,再次进行一致性测验。结果仍然制成一个九分量表。合并两个九分量表,我们会发现,最终的这个九分量表中,每一等级包含四个命题陈述。

“对偶比较测量”的矫正

前面，我们从理论上探讨了“对偶比较测量”的研究方法。然而，在该研究方法的设计中，有几处任意性较大，这是读者应当认识到的，前提是读者愿意修正研究的任意性，达到自己的研究目的。首先，研究中采用的是九分量表，九分量表的选择就很明显是任意的，因为只要量表适于我们的研究，就可以选择它（这一点对Q检验来说也是正确的）。然而，在这一点上存在有一些共识，那就是九分量表代表了概念区分的最佳数目，十五分量表可能划分太细，而对于许多研究来说，三分量表可能不允许进行充分的等级划分。进一步说，制定一个标准量表是有益的，如果必要的话，某一次测量的结果可以和另外的测量结果进行比较。稍后我们将讨论这个问题。

另一方面，有可能出现这种情况，需要运用大的或小的量表，这时，有没有什么便利的方法，可以帮助研究者选择最合适的量表？要回答这个问题，我们可以回过头去看先前我们关于指标的可辨别性的讨论。一般来说，假定我们随机选择的、用来测试一致性指标的数目是n，对这n个命题陈述进行对偶比较研究，其中概念不清的命题陈述的数量属于可以接受的范围，那么量表就可以至少包括n个等级。因此，如果在18个命题陈述中，比如说，有2个命题陈述的概念纠缠不清，我们就可以使用十八分量表。然而，这种方法只能确定量表等级的上限，研究者必须在这个范围内确定量表的长度。

在断定某个评分者的前后一致性时，我们从2 000个命题陈述中随机抽取了18个命题陈述构成样本。这仍然是个任意性决定。进行一致性测试的样本可大可小。然而，抽取的命题陈述的数量越多，研究所允许存在的定义不清的命题陈述数也就越多。或者说，为了说明这个细微的差别，样本中所包含的命题陈述数越多，各命题陈述所得票数构成完美等级的偶然性就越小。另一方面，再看看n个命题陈述可能包含的配对数的计算公式，也可以发现，命题陈述的数量越多，产生的配对也就越多。因此，样本数量超过25，在实际中就不可行了，因为根据公式，$n=25$，配对数将是300对，这个数字以及随之产生的需要进行的判断，已超出了任何一个评分者的能力范围。本测验的样本容量应当保持在10～25之间。我们抽取

了 18 个命题陈述,也是为了减少从十八分量表向九分量表合并过程中的工作量。

“对偶比较”测量的说明

我们来讨论“对偶比较测量”的结果。实验进行到这一步,2 000个命题陈述都有了测量值,大小从 1 到 9 不等。必须注意,这些测量值是相对数值,就是说,这些数值仅仅反映了每一命题陈述在整个 2 000 个命题陈述中的相对位置。这种情况在“对偶比较”测量中表现不是很明显,而在 Q 检验中却很清晰。因为,在 Q 检验中,我们是将每一指标与其他指标进行比较从而得出测量值的,所以很明确,测量的结果反映了某一命题陈述相对于其他命题陈述的位置。而在“对偶比较”测量中,我们是将样本中的 18 个(或 36 个)命题陈述作为媒介,通过它们对所有命题陈述进行相对测量的。因为样本是从所有命题陈述中随机抽取的,所以有理由认为样本代表了总体。当我们合并命题陈述将之制为九分量表,并使用两次抽取的样本对剩下的 1 982 个命题陈述进行等级测量时,我们所得的数据就是全部命题陈述的相对测量值。有些研究中,与“相对测量值”相对,需要的是“绝对测量值”,或“真测量值”。例如,我们可能希望在“理想型态”下测量一个变量。“对偶比较测量”很容易修订,可以用来进行“绝对”测量。在这些例子中,一致性测验的结果不能用来确定量表的等级数。此时,考虑到人们期望的绝对或“理性型”,我们可以建构一个样本组,使用这种样本组,也可以完成对所有命题陈述的测量。

其他讨论事项

讨论“对偶比较”测量时,我们没有讨论掩蔽问题。因为,如果我们对一种不是正态分布的数据进行 Q 检验,我们就会问这样的问题:在 Q 检验中,是否有必要掩蔽每一个命题陈述?这个问题到现在仍然没有确定的答案。对概念进行掩蔽,与不对之进行掩蔽相比,测量的结果是否会存在显著的差异,关于这个问题,人们至今还没有进行验证。然而,一种可能情况是,随机选择两个或两个以上的评分者,计算他们之间的信度系数,也许可以看出因为没有掩蔽命题陈述而产生的偏见。此时,研究者必须做出决定,是否掩蔽他

的命题陈述。

最后一个注意事项是：假定如上所述，用"对偶比较"测量方法来研究上面从一场危机中选出的2 000个命题陈述。这2 000个命题陈述中，每一个命题陈述都有一个相对测量值。进一步假设，过去我们曾做过同样的测量，是对从另一场危机中选出的2 000个命题陈述进行过相对测量。现在，我们想对这两场危机进行比较。因为，测量是在相对的层面上进行的，与我们讨论Q检验时出现的情况相类似，我们不能对两组独立的测量值进行合理的比较。

一个解决办法是，合并这两组命题陈述，按照"对偶比较"的程序，对全部4 000个命题陈述进行重新测量。当然也确实存在有相对简单的方法，用在Q检验中，或者稍加修改，用在"对偶比较"测量中。

可以从两场危机中各随机选择9个命题陈述组成一个样本，做好标签，从一场危机中抽取的命题陈述标上A，从另一场危机中抽取的命题陈述标上B。接下来，将这18个命题陈述放在一起，打乱原来的顺序，重新排序。在一张白纸上，列出从危机A和危机B中抽取的命题陈述，共18个，写出其排序号，将先前的测量值对应地写在每一命题陈述的旁边（即，它们分属于危机A和危机B时的相对测量值）。例如：

关于A的		关于B的	
1	7	3	2
2	6	4	5
7	3	5	7
8	1	9	9
11	9	13	6
……		……	

仍然使用标签A和B，保证每一命题陈述的序列号是唯一的验证记号。用同样的程序进行测量，即，将18个命题陈述进行配对，对每一对命题陈述进行比较。将结果合并成九分量表，这样，每一命题陈述就有了一个相对测量值。通过这种方法，可以对每个命题陈述的旧的或原先的测量值与新得出的测量值进行比较。这种比较可以说明，如果两组命题陈述一起测量的话，如何对原来的测量值进行修订。例如，比较可以显示出，危机A中的每一个命题陈述

必须减少降低一个测量值,而危机 B 中的每一命题陈述则必须上升两个测量值。那么,修订原来的测量值,对两场危机进行比较,通过这种方法就相对简单多了①。

本章小结

用来测量文献资料表达强度的方法有三种,本章介绍了其中的第二种。只有当 Q 检验在某一项研究中不可行,或者说对于所有的研究设计来说,Q 检验不是最有利的方法时,才可以选择其他的方法来代替它。本章中,我们还介绍了"对偶比较"的测量方法,这种方法是一种相对测量方法,它不考虑数据的正态分布状况如何,仅关注那些有助于对众多数据所得的结果进行比较分析。

① 很明显,该程序存在着一些假设,不论是用于 Q 检验中,还是将之修订以用于对偶比较测量中。在本例中,该程序假定,如果主题 A—3 比主题 A—1 的测量值高的话,使用危机 B 中的主题对这些主题进行重复测量,并不能改变"A—3 的测量值高于 A—1"这个关系。也就是说,不论赋予 A—3 和 A—1 什么样的新的测量值,A—1 的新测量值将高于或等于原有的测量值。

第 6 章[1]

报刊主题分析:竞选总统与“赛马”

本章集中分析了 1988、1992 年两年的总统大选期间报刊媒体有关竞选政治内容的报道,尤其是那些著名报纸相关议题的“赛马场风云(horserace)”式报道。虽然在报道 1992 年的总统竞选中,大众媒介希望能够在报道国家重大议题方面做得更好,但是很明显,它们自身不能抵制这种赛马式报道的诱惑和偏好。

社会公众如果以报纸新闻报道为视角来看待 1988 年那场他置身其中的美国总统竞选,可能会得出这样的结论:这场总统竞选,总与乌七八糟的政治花边新闻搅和在一起。譬如,有关在武装坦克上参加游行的民主党总统候选人的政治讽刺速写、监狱里犯人被批准休假的危险性以及绘声绘色地描述某政要人物在造访一家旗帜制造厂时被织物罩住险些被卷入机器中,等等,诸如此类新闻由头报道与总统竞选报道交织在一起。1988 年有关总统选举的很多报道集中在候选人提名、喧嚣的竞选活动场面以及新政府公职的竞争方面。社会公众很少能够听到关于未来政府的重大政策议题,以及如果候选人当选总统将会采取何种措施解决这些问题的谈论和分析。

一些职业政治评论家竭力地提醒公众注意,在 1980 年到 1990 年期间,随着报道轰动事件(staged events)和“谁跑在先(who's ahead)”这一主题,美国的大众媒介正在经历着一场积极的转变。他们认为,在现代美国的政治进程中,美国的媒体似乎从来没有像在 1992 年的总统竞选中表现出来的那样具有关注国家的自我

---

① 本文使用已获得两位作者马克·米勒(Mark Miller)和布里安·德纳姆(Bryan Denhham)的许可。

意识。

> "随着 1992 年竞选运动的开始",洛杉矶时报的媒体专栏评论家托马斯·罗森斯蒂尔说,"媒体吵嚷着要做的事情各不相同。简而言之,印刷媒体和大众广播的从业者受到了对他们批评的影响。这些批评指责媒体在 1988 年竞选期间,没有给予这个国家所面临的最为紧迫的问题以足够关注……"

在 1988 年竞选期间以及 1988 年以后,新闻传媒界专职的研究人员和学术界的研究人员分析和批判了美国报纸对民意测验的滥用。这些研究者指出,民意测验报告,正像通常的媒介报道一样,仅仅将主要精力集中于那些轰动事件和乔治·布什与麦克·杜卡基斯之间竞选的"赛马场式的风云",而漠视或者根本就不关注那些维系民生的重大社会公共问题。

艾尔灵·克里斯皮解释道,如果能够适当采用,民意测验能够促进社会对公共事务的认知。阿尔波特·肯特里尔也探讨了民意测验的运用问题,断言"如果能够合理运用,负责任地报告测验成果,并且不经过持这样或那样观点的鼓吹者的中间环节,则民意测验就可在把公共议程转化为政治过程的过程中,起到巨大的作用。"

但是,在 1988 年,通过民意测验获得的信息看起来又一次仅仅是集中关注了公共部门竞选的政治过程的细节和表面的问题。一个资深的评论家已经详细讨论了最近几十年出现的这种现象。

这一研究考察了对 1988 年和 1992 年在报纸上出现的关于政治议题和竞选主题的探讨。并且,该研究提出了议题与相关术语(issue terms)的内在不一致性,而且考察了其客观的或映像的含义,以此为基础,力图实现我们所指的议题(issue)的概念化过程。这样,虽然财政预算和婚姻忠诚都是一场选举运动的议题,但是在新闻报道所提供的信息量方面,这些问题却没有受到同样的重视。

在一个完整的政府民主体制中,新闻从业者(journalist)扮演着不可或缺的作用。在选举运动期间,这些人来区分哪些是政策问题,哪些是候选人所造成的映像(candidate image)。公众喜好的问题,如候选人的性生活问题等,与医疗保险和社会安全议题存在着基本的区别,而这些问题又与人们的生活密切相关。这样说不是要贬低候选人的个人品质的重要性,而是为了说明,为了娱乐的目的而挖掘一些问题时,新闻记者可能不能完成他们所担负的向美国选民传递政策议题信息的责任。

托马斯·帕特森探讨了政策(policy)和竞选(campaign)之间的区别,他注意到在竞选过程中,新闻被某些具有明显的非此即彼的事件所主导。总而言之,这是暗喻,大众媒介把大量的注意力放在了那些能够明显区分出选举人的议题上,这些议题能够激发竞争冲突,并能用简单的话语来表示(如堕胎和枪支控制等)。

但当报纸报道这些应当经过修饰的、通常具有轰动效应的问题,而忽略了这个社会所面临的重大问题时,美国的选民能够更好地被大众媒介告知吗?或者,媒介对于那些有轰动效应的、能够激发大众好奇心的事件的报道,是否使得公众对于重大政治议题的忽视更为严重?

1988年,《洛杉矶时报》的媒体批评家大卫·肖的研究指出,在通常情况下,人们为大量的关于候选人本身的报道信息所淹没和胁迫,但是由于人们对于整个公共政策信息缺乏了解而对社会重大议题茫然无知。罗伯特·恩特曼接下来的研究也表明了这点。他提出了一个矛盾的命题陈述"非增长的充足(abundance without growth)",来说明现代大众媒介报道的困境:在竞选期间,各种信息铺天盖地扑面而来的同时,人们却对真正关系到自身利益的政治交易茫然无知,并对社会议题缺乏了解。

在肖的论文里,亚利桑纳前州长布卢斯·巴比特把大众媒介关注竞选策略和候选人个性品格,而不是社会重大议题的原因归结为不断增加的民意测验和其他一些因素。肖解释道,民意测验把候选人区分为胜利者和失败者,为选民提供了一种相对来讲可以触摸到的手段来了解竞选,这也激发了人们对竞选的兴趣。事实上,一些人已经指出,媒介报道的竞选主题是一个天天面对的新闻魔鬼,通过每天报道民意测验的领先与否,刺激那些对于选举漠不关心的人给予大选更多的关注。然而,可以合理推测,即使是参与了一个人山人海的竞选政治推销大会,结果选民的脑子里仍然还是对现实中重大社会问题的了解知之甚少。那些报刊在竞选期间真正究竟向广大选民"猛劲地忽悠"些什么样的竞选主题呢?

斯科特·莱特茨安的见地颇为独到:

> 虽然人们还在就民意测验的价值及其对政治过程的贡献争论个不休,但有一个事实,无论赞成者还是反对者都会同意,即这种调查对那些对投票漠不关心的人造成了一种挥之不去的影响。民意测验使得竞选运动定型化(freeze-frame)。当候选人把注意力的焦点集中于民意测验的良好表现而不是直截

了当地处理和直面与美国人民密切相关的政策议题时,议程的设置便发生了偏向,拐了个弯。

本研究的目标意在考查1988年和1992年两次大选期间,重要的民意测验所关注的中心议题究竟是什么。如果在1992年,那些大牌宣传媒体如它们所宣扬的那样大有作为,那么我们就应当较少地观察到媒介报道的"赛马场式的风云"以及一些能够把候选人清楚区别并且引起冲突的表面映像议题,而是从社会现实的根本需要上更多地看到需要政府采取政策解决的重大议题。

本研究也考查了竞选阶段若干报纸报道的区别。考查竞选阶段非常重要,原因在于每个阶段的新闻主题都各不相同。换言之,在竞选的早期阶段,所有的候选人都讨论他们在某些特定议题上的观点,试图把这些看法能够传达到潜在的投票者。在这个阶段,这场赛马还没有获得很多公众的喝彩。在两党各自的参选鼓动大会和选举大会后的时期,各党的候选人都努力成为这场总统赛马的领先者。本研究通过对三种报纸分类考查编码分析,进而探讨了选举报道的基础。即,报纸报道是否为选举过程所左右?如果是这样,三种报纸的报道将大致类似,或者,报纸的报道是否更多地依赖于某些新闻从业人员的判断力?

方　法

本研究的数据来源于权威的报业数据库——莱尼克斯(Nexis)数据库,数据分析所采用的计算机软件是VBPro。从莱尼克斯数据库里下载了来自《洛杉矶时报》、《纽约时报》和《华盛顿邮报》的文章,并进行了甄别、格式化以进行按照字母顺序的排列出词频表。选择这三种报纸进行分析的原因是,每种报纸都在其新闻从业水平方面具有很高的知名度,而且它们都经常进行民意测验并把结果公之于众。因此,这三种报纸可能可以代表民意测验的专业操作水平的最高水准。

所选文章来自于新闻报道,不包括社论、专栏文章、读者来信等,并且以百分比或分数的形式报道民意测验的结果。民意测验要基于随机概率抽样样本;不考虑会议代表(convention delegate)和投票后民意调查(exit polls)。删去了表格,把似是而非的参考题压缩转换为关键词。

经过以上筛选过程,选出468篇样本文章,包括了1988年和

1992前一年的总统竞选期间所刊出的报道。这些文章中，有90篇来自于《洛杉矶时报》，237篇来自于《纽约时报》，141篇来自于《华盛顿邮报》。在1992选举年，《洛杉矶时报》主要依赖于投票后的民意调查，所以样本数量少于另外两种报纸。

竞选阶段

考查了竞选阶段（两党参选鼓动的大会前，大会期间，大会后），再加上竞选年专刊，以及三种报纸。探讨这些竞选阶段，可以使我们了解在各个阶段媒介的行为，以此确定"赛马式"的报道在竞选总统的政治活动中其新闻含量是增加了，还是减少了。

选取关键词

考虑到"赛马式报道"的特点和新闻出现频率，发现有23个关键词与候选人在竞选过程中的相对位置密切相关；没有选择的42个词，由于模棱两可程度各不相同，并可能转移内在分析的确定性，所以未选择（例如，"领导"一词包括在短语"民众期望的领导"和"健康计划的领导"。

镜像索引包含了14个词，索引范围包括非此即彼分类的词（如堕胎、枪支控制和死刑）、激发矛盾冲突的词（如犯罪和毒品）以及品格描述的词（如婚姻忠诚和伊朗反对派）。正如托马斯·帕特森和多里斯·格莱博的研究一样。映像议题给予表面的政治立场和个人品格更多的关注，而给予特别的政府政策的关注很少。例如，一个候选人是否赞同北美自由贸易协定这个问题与一个候选人是否也有墨西哥裔的美国孙子这一问题即存在略微的差别。同样的，是否一个候选人赞同在私立学校实行学校教育券制与一个候选人是否恰好是领过洛克菲勒奖学金的研究生也不同。在竞选过程中媒介报道应当考查候选人的品格，这一点为人们所赞同，但是不能忽视的一点是，媒介也应该为公众提供那些重要的公共事务的相关信息。

重大问题索引共包含了24个词，每个词语都聚焦于政府的政策可行性（如预算、国防和医疗保险）。这些词语不包含非此即彼的分类，也不聚焦于候选人的个人品质，而主要聚焦于内政和外交政策议题。

通过编码分析，可以帮助我们发现在1988年和1992年的总统竞选过程中，赛马式的报道和议题关键词占每篇文章的比例是否存

在显著的差异。由于各篇文章长短不同,故这里采用了百分比的方式来计量,而不是采用频率的方式。

结　果

本研究采取的是三个维度的向量分析,即对大选年的影响、竞选阶段和三种报纸的每篇文章中赛马式报道和议题报道的百分比进行分析(见表6.1)。

表6.1　在竞选阶段、大选年和每种报纸的所选择样本文章中,赛马式报道、重大议题报道和映像报道的平均百分比(percent)

	竞选过程					
	大会前		大会中		大会后	
	1988	1992	1988	1992	1988	1992
赛马式报道						
洛杉矶时报	0.75	0.56	0.81	0.98	1.24	1.10
纽约时报	0.63	0.42	0.39	0.81	0.65	0.82
华盛顿邮报	0.81	0.41	0.52	0.49	0.76	0.68
重大议题报道						
洛杉矶时报	0.26	0.90	0.31	0.45	1.02	1.00
纽约时报	0.73	1.09	0.45	0.89	0.95	0.53
华盛顿邮报	0.55	0.94	0.85	0.50	0.89	0.87
映像报道						
洛杉矶时报	0.03	0.06	0.11	0.11	0.51	0.09
纽约时报	0.18	0.23	0.06	0.41	0.19	0.03
华盛顿邮报	0.08	0.07	0.32	0.02	0.22	0.03

依据初步分析,发现在大选年和竞选阶段之间存在着显著的交互关系($N = 468$, $F = 3.512$, $DF = 2$, $P < 0.05$)。在两党大会阶段,赛马式的词语在1992年的开始阶段较低,但是高于1988年同期水平。而且,在两党会后阶段,1992年这一比例也高于1988年。

这表明，随着竞选的逐步发展，1992 年的赛马式民意测验报道在增长。虽然在报道 1992 年的总统竞选中，大众媒介希望能够在报道国家重大议题方面做得更好，但是它们自身却不能抵制这种赛马式的报道。

在对赛马式报道的均值计算后，发现均值间存在显著差异（$N = 468$，$F = 3.074$，$DF = 5$，$P < 0.01$）。三种报纸之间报道的差异的原因在于，报社在选择报道竞选内容方面的编辑策略存在差异（$N = 468$，$F = 3.281$，$DF = 2$，$P = 0.01$）。这种差异就意味着，在报道竞选的过程中，报纸的编辑来决定要报道竞选过程的哪些方面。表 6.1 还包括每种报纸在竞选年和竞选的三个阶段的平均百分比数据。

在竞选进行阶段，也存在显著性的差异（$F = 4.097$，$DF = 2$，$N = 468$，$P = 0.017$）。在 1988 年，两党参选鼓动大会前报纸的每篇文章采取赛马式报道中的关键词的均值是 0.68；在 1992 年，这一数值降低至 0.46。在两党大会阶段，这一均值是 0.54；然而，在 1992 年，这一均值上升到 0.71，因此，这也再次说明了之前讨论过的交互关系。在两大党派（民主党、共和党）参选鼓动大会后的相当长一段时期，这一均值基本上保持了不变。

重大议题报道的关键词包括那些与政府政策密切关联的词。当把两个大选年的数据都考虑进来，发现《纽约时报》的重大议题报道的数量高于《洛杉矶时报》和《华盛顿邮报》。然而，这种差异值没有超过均方差值，所以差异不显著。

总的来看，这三种知名报纸在 1992 年的报道中，比 1988 年更多地报道了重大社会议题。然而，应当注意，在两党各自参选鼓动大会前阶段，这种差异最大，而本研究所选取的样本也主要来自于这一阶段。在两党各自参选鼓动大会前，报纸报道重大社会议题的关键词数量远高于 1988 年的数量。与赛马竞技模式相反，伴随着两党各自参选鼓动大会的三个阶段，重大社会议题报道的数量逐渐递减，最终低于 1988 年的两党各自参选鼓动大会后阶段报道的关键词数量。我们还可以认为，在报道 1992 年的总统大选的初期阶段，报纸的从业人员决定报道更多的重大社会议题，但是，最终，他们不能抵制报道这场总统竞选赛马场面中的热热闹闹的表面问题。

当我们把三种报纸分开来考查时，发现《纽约时报》运用了 14 次映像报道的关键词，比《洛杉矶时报》和《华盛顿邮报》的多。至于重大议题报道，三种报纸之间并不存在显著性差异。

虽然在统计意义上并不显著,但《洛杉矶时报》和《华盛顿邮报》在使用映像词汇报道的数量呈略微下降趋势。在对赛马式报道进行计量统计分析时发现,大选年和竞选阶段之间存在显著交互影响关系。镜像报道在1992年竞选初期阶段比较低,但是也像赛马式报道一样,在两党各自参选鼓动大会阶段开始超过1988年的数值,之后又降低很多。

讨 论

在关于1992年总统竞选报道的争论中,本研究提供了一些支持性的论据,即发现在1992年,赛马式的民意测验报道主题相比以前有所降低,并发现了报纸主题报道的一些发展趋势。但是,本研究也发现,虽然新闻从业人员本意是想在报道重大社会议题方面做得更好、更有进步、更获得社会公众的普遍认可。但是在面对新闻由头强健的总统竞选题材诱惑时,三种知名报纸却还是花费大量的功夫循规蹈矩来专门报道总统竞选那种"赛马场式"的竞选竞技场面了。

三种报纸的赛马式报道之间存在差异说明,在选择报道哪些问题时,新闻从业人员实际上还是在起决定作用。如果没有发现在三种报纸的报道之间存在差异,则我们便可认为大选本身就决定了新闻报道的风格。这再次证明了这种观点,即在报道的过程中,新闻从业人员可以"有所作为(do things differently)",并且还是可以大有作为的。

一般来讲,日常的"谁跑在先"这类报道必然会对选民的明智判断有所误导和帮倒忙,添乱,竞选过程中的花里胡哨的报道也不可能真正做到对于选民需要的政治价值的积极性倡导。如果大众媒介能够在未来的1996年总统大选过程中,更好地改进其有关重大社会议题的报道,而不是一味迎合新闻鳌头去挖掘那些与候选人密切关联的、作为人们茶余饭后笑谈品资的"赛马场式的竞选报道",那么美利坚合众国的广大选民们方有望可能在未来比较真切地了解和感悟到这个国家所真正面对的必须解决的重大社会现实问题,因为这些现实的问题才是与每个人的生活最为密切相关的。

第 7 章[①]

教育政策过程评价：群众来信关于教育政策的话语分析

研究背景：信访如潮涌动

《中国教育年鉴》记载了教育部历年来办理群众来信来访的情况，本文作者仅仅摘取其中从 1997 年到 2001 年期间的信访记录枚举如下（表 7.1）。

从上述记录的情况可以看出，信访与上访相互关联，信访数量与教育政策调整所涉及的群众利益范围密切相关，信访的特征以及信访反映的突出问题与教育政策的执行过程相互关联。

教育政策的制定、宣传、执行、评估等过程，都是通过相关的话语形式来表达的。公众自发地向政府寄信，反映对有关教育政策的看法，围绕教育政策与自身利益之间的关系，表达一些朴素的认识和直接的愿望，这是在我国法制社会环境下民主参与决策的途径之一。同时，信访从另一个侧面反映出教育政策在基层的作用和效果表现。

本章选取了 2000 年和 2002 年教育行政部门受理的 225 封反映教育政策问题的公众来信，试用社会语言学和统计学的方法对来信话语特征进行分析。初步揭示了公众来信的话语秩序、问题域结构、话语谱系等基本特点，以及与相关的学术话语之间存在的主要差异，从而对公众认可的热点问题与教育政策之间的关系，进行话

---

① 本文写作过程中得到中国国家教育部教育发展研究中心主任张力教授的悉心指导，在此表示感谢。本文的文责仅由作者李钢博士本人承担。

语意义上的探索。

表 7.1 1997—2001 教育部办理信访情况一览表

历年信访基本情况	信访特点和信访的突出问题
2001 年:教育部信访局共处理群众来信(来电)21 000 多件,接待群众来访 4 100 多人,其中集体上访 76 批,约计 590 人	一些教育热点、难点问题相对集中,涉及群众权益方面的联名信访和集体上访多,电话信访大幅度增加。信访的突出问题是:民办教师被辞退后的安置问题,教育乱收费问题,中小学教师工资拖欠问题,高校招生问题,举报干部违纪问题,职业教育、社会力量办学、成人教育问题等
2000 年:教育部信访办共处理群众来信(电)21 900 多件,接待群众来访 4 750 多人,其中集体上访 60 多批,约计 530 人	群众对全面实施素质教育和教育管理体制改革反响强烈,参与意识强,对采取政策减轻中小学生课业负担普遍表示赞成。信访的突出问题有:民办教师的安置问题,教育乱收费问题,高校招生中的不正之风问题,学生权益受到损害举报,拖欠教师工资问题,基层干部违纪举报
1999 年:教育部信访办共处理群众来信(电)2 300多件,接待群众来访 4 200 多人,其中集体上访 40 多批,约计 300 多人	社会对高校扩招、筒子楼改造工程和民办教师转正政策给予热烈拥护。信访反映突出的问题有:教育乱集资乱收费问题,中小学生课业负担过重,校园周边环境安全,教师队伍用人腐败问题,教师职称评定问题,大学生毕业分配收费问题,社会力量办学不规范和欺骗性招生问题,教师体罚学生问题,各类干部违纪问题举报
1998 年:教育部信访办共处理群众来信(电)21 200 多件,接待群众来访 3 960 人,其中集体上访 82 批,约计 410 人	信访反映突出的问题有:教育乱集资乱收费、各种形式的"捐资助学"问题,中小学生课业负担过重,拖欠教师工资,校园安全问题,师资队伍建设中的问题,高校与地方合办研究生班质量低的问题,社会力量办学不规范问题,基层干部违纪举报
1997 年:教育部信访办共处理群众来信(电)18 222 件,接待群众来访 3 910 多人,其中集体上访 29 批,约计 310 人	信访突出反映的问题有:教育乱集资乱收费问题,拖欠教师工资问题,举报揭发学校财务管理腐败问题

注:以上记录均摘录自《中国教育年鉴》,分别来自:2002 年卷:131;2001 年卷:104,105;2000 年卷:128;1999 年卷:161;1998 年卷:120

公众书信的基本情况以及信访较为集中的主题

“话语分析（discourse analysis）”是对话语和语篇进行分析，揭示话语之间内涵的话语意识，通过特定的语境关系分析话语当中所关联的其他社会意识形态关系。20 世纪 70 年代，话语分析（或语篇分析）开始成为语言学的独立领域和社会语言学的分析方法。同时，受到了社会科学研究界的注意，并被拓展应用到公共行政与公共管理等方面的研究①。在后现代教育理论中，话语分析由于能够消解语言形式对于交往意识的“遮蔽”，增进教学双方交流互动的有效沟通。因此，话语分析方法也被尝试引入教育组织的语言行为分析之中（史密斯，2000：199-206）。

话语分析所需的基础材料是十分广泛的，而此处所采用的仅限于公众来信，这些来信的主要特点有：一是属于自发性质的社会话语，是一定时期社会制度环境中的产物；二是具有明确的自我表述意识以及相关的话题表达形式；三是使用了汉语规范的书写方式，不存在不同语言符号之间的意义转换问题；四是所表述的价值观与国家政治和法制相关的意识形态价值观基本一致；五是写信人与信件受理者之间存在着确定的关系②。按照以上的假设，我们确定了对信件内容进行分析的技术路线。

首先，我们选取 2000 年和 2002 年教育部有关部门收到的 225 封公众来信，作为一个语料整体进行分析。经过甄别，所有样本来信均系反映现行教育政策的执行情况、存在问题和改进意见建议，没有信件牵涉解决个人（群体）实际问题或举报违法乱纪案件等内容。其次，参照话语分析方法和预设指标，对样本信件进行编码，按照信件的特征设定的编码分别为：①信件受理机关；②寄信者所在省（自治区、直辖市）；③寄信者所在地类别；④寄信者社会身份；⑤信件收到日期；⑥信件表达的主题内容；⑦信件主题引用的信息来源。

---

① 经网络主题词检索，作者发现在教育领域中的“话语分析”（或“语篇分析”）以及有关话语分析的论文或专著，绝大多数与语言（含外国语）教育教学或语言学研究等相关。在教育政策分析及政治学、社会学等领域，只有少量研究成果。如：Corson，1995；Ockwell，2001；麦克斯怀特，2002；福克斯，米勒，2002 。

② 关于以上假设的理论背景，请参考：刘润清，2000：287；斯考伦等，2000：97；桂诗春，宁春岩，1997。

然后,我们将样本信件的主要内容归纳为主题词,并就主题词频次及其他数据进行统计分析。由于公众来信的写作风格多系口语化的议论文体,信件的语篇结构呈现出灵活多样且随意性强的特点,我们在对信件表达的主要内容进行主题词归类时,尽可能选用出现频次较多的、约定俗成的、贴近报刊媒体的正式用词,这样做的目的是为了增强统计数据的集中度。避免在话语理解方面产生过大的歧义。最后,如果统计结果的可靠性没有通过检验,则须调整预设指标的范围与相应的编码系统。如果统计结果通过检验,可进行下一步的综合性研究。以下是我们对来信的分析策略图式(图7.1),以及用这种图式分析得出的不同地区公众来信频数的若干特点分布一览表(表7.2):

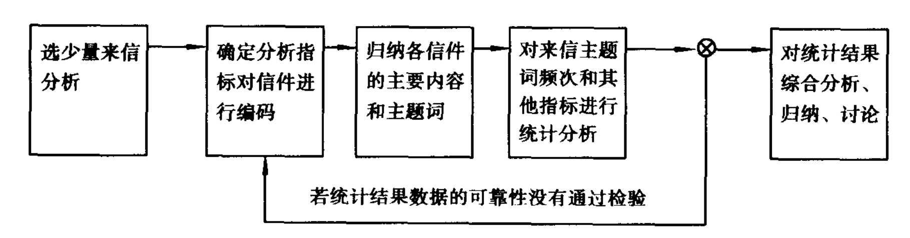


图7.1　群众来信主要内容话语分析的技术路线图示

表7.2　不同省(自治区、直辖市)的公众来信频数及其若干背景指标

	A.2001年人均GDP/元	B.2000年人均受教育程度/年	C.2000年来信频数	D.2002年来信频数	E.来信频数总计	占E百分比
江苏	12 922	7.85	18	15	33	14.7%
山东	10 465	7.58	17	9	26	11.6%
浙江	14 655	7.46	16	4	20	8.9%
湖南	6 054	7.8	12	6	18	8.0%
北京	25 523	9.99	10	5	15	6.7%
上海	37 382	9.3	5	4	9	4.0%
云南	4 866	6.33	7	2	9	4.0%
安徽	5 221	6.98	7	1	8	3.6%
广东	13 730	8.07	3	5	8	3.6%

续表

	A. 2001年人均GDP/元	B. 2000年人均受教育程度/年	C. 2000年来信频数	D. 2002年来信频数	E. 来信频数总计	占E百分比
黑龙江	9 349	8.25	5	2	7	3.1%
四川	5 250	7.06	5	2	7	3.1%
贵州	2 895	6.15	5	2	7	3.1%
陕西	5 024	7.71	6	1	7	3.1%
山西	5 460	8.02	4	2	6	2.7%
福建	12 362	7.49	6	0	6	2.7%
天津	20 154	8.99	1	4	5	2.2%
河北	8 362	7.74	4	1	5	2.2%
江西	5 221	7.55	2	3	5	2.2%
吉林	7 640	8.24	3	1	4	1.8%
河南	5 924	7.72	2	2	4	1.8%
湖北	7 813	7.77	4	0	4	1.8%
广西	4 668	7.57	3	1	4	1.8%
内蒙古	6 463	7.76	3	0	3	1.3%
辽宁	12 041	8.41	3	0	3	1.3%
海南	7 135	7.68	1	0	1	0.4%
新疆	7 913	7.73	1	0	1	0.4%
重庆	5 654	7.28	0	0	0	0.0%
西藏	5 307	3.43	0	0	0	0.0%
甘肃	4 163	6.54	0	0	0	0.0%
青海	5 735	6.12	0	0	0	0.0%
宁夏	5 340	7.03	0	0	0	0.0%
合　计	7 543	7.62	153	72	225	100.0%

资料来源：A列引自《中国统计年鉴·2002》，B列据2000年第五次人口普查数据测算，其余由作者通过实际调查自行统计获取。

来信话语发生的地域特征分析

不同年份公众来信的地区发生率具有十分稳定的一致性,但尚未证实其与地区经济文化水平之间存在显著的线性关联。

在225份样本信件中,2000年共有153封,2002年10月以前有72封,分别来自26个省(自治区、直辖市)。我们对表7.2所示2000年与2002年不同地区的来信频数(C列与D列)进行相关分析,得到相关系数为0.7919,通过了1%的显著性检验,这表明两年的公众来信在地区来源上保持着十分稳定的一致性。

从两年的来信频数总计看,公众来信列前5位的地区是:江苏(14.7%)、山东(11.6%)、浙江(8.9%)、湖南(8%)和北京(6.7%),这5个省(市)的来信合计占总量49.8%。并列第6位的是上海和云南(均为4%),并列第7位的是安徽和广东(均为3.6%),并列第8位的是黑龙江、四川、贵州和陕西(均为3.1%),以上前8位共13个省市的来信合计占全国总量的77.4%。

教育承载着知识信息的传授、积累、整合、创新和传播等多方面的功能,按照文化信息传播理论推知,公众来信发生率应当与当地的知识信息交流水平之间有一定的关系。例如,表7.3关于寄信者所在地的统计表明,直辖市、省会和自治区首府城市和地级市的公众来信较多,2000年占总量的73.2%,2002年略有下降,仍然占62.5%,两年的合计占69.8%;县镇和农村的来信比例有所增加,从2000年的22.9%上升到34.1%。两年的合计占26.9%。显然,来信频数反映的公众参与程度,与所在地的经济发展、教育发展,尤其是信息交流等综合条件有比较密切的关系。

表7.3 不同地域的公众来信频数对比

寄信者所在地	2000年人均受教育程度/年	2000年		2002年		总计	
		来信频数	占比例	来信频数	占比例	来信频数	占比例
直辖市、省会(自治区首府)	~10	43	28.1%	29	40.3%	72	32.0%
地级市	~9.5	69	45.1%	16	22.2%	85	37.8%
县 镇	8.73	13	8.5%	10	13.9%	23	10.2%
农 村	6.85	22	14.4%	16	22.2%	38	16.9%
地域不详	—	6	3.9%	1	1.4%	7	3.1%
合 计		153	100.0%	72	100.0%	225	100.0%

注:2000年人口受教育程度数据引自《从人口大国迈向人力资源强国》。其中,直辖市和省会(首府)和地级市的数据,系据均值粗略估算,仅作参考。

那么，能否可以推论，越是经济发达或者教育水平相对发达的地区，越可能是来信反映政策问题较活跃的地区呢？为了验证这一关系，我们对表7.2中A列各地区的人均国内生产总值(GDP，一般用于衡量经济发展水平)和B列的人均受教育年限(一般用于反映教育发展的综合水平)进行相关分析，相关系数为0.6246(通过1%的显著性统计检验)，由此可以看出，地区的经济发达水平与教育发展水平存在着较好的正相关关系，这在近年来的其他分析中曾被反复证明(张力，1998：83-93；国家教育发展研究中心，2000：45-48)。然而，我们用表7.2的A、B两列与E列的两年来信频数总计进行相关分析，其相关系数却均未通过5%的显著性检验。同样，用表7.3的不同类型城市、县镇和农村的人口平均受教育程度与其两年公众来信频数进行相关分析，其相关系数也没能通过5%的检验。也就是说，某省(或者城市)的经济发展水平和教育水平越高，该省(或者城市)反映教育政策问题的来信并非越多。

这可以解释，在来信频数排列前8位的省市之间，经济及教育发展水平的差异还是较大的，同时，也可以推测，一些省市的公众来信占全国的比例很高，或者农村来信比例的上升幅度比城市要大，而重庆、西藏、甘肃、青海、宁夏等省区则暂时没有来信，可能有经济和教育以外的原因。因为直接向国家级政府写信来表达意见和建议毕竟只是诸多渠道中的一种，所以，只有在当地的文化传统和社会环境中，具有凝聚有民诉的文化意识，才会有可能形成(或出现)这种当地群众愿意采取的诉求方式。此外，还须考虑到由于样本较少，所导致的偶然性因素的较大影响。

哪些人构成了自发写信的基本群体

不同界别写信者的信件份数和主题频数的差异较大，写信的基本群体是教师，其中中学教师的参与程度最高。

按照社会语言学的观点，由于不同人群来自不同的社会阶层，他们的社会化适应程度、持有的价值观与选用的话语形式和表达方式等都存在着差异，不同的人群在针对同一类教育话题所表述的话语时，带有特定的话语基调和明确的话语特征(斯考伦等，2000：155-170)，这些特征有助于识别教育问题所体现的社会阶层的意识与利益。中国社会科学院的《当代中国社会阶层研究报告》以组织资源、经济资源、文化资源占有状况作为划分社会阶层的标准，把当

今中国的群体划分为十个阶层,即国家与社会管理阶层,经理阶层,私营企业主阶层,专业技术人员阶层,办事人员阶层,个体工商户阶层,商业服务人员阶层,产业工人阶层,农业劳动者阶层,城乡无业、失业和半失业人员阶层(陆学艺,2002:8-23)。我们从写信者自己标明的身份出发,对他们的身份状况进行了分类统计,其中,将人大代表、高中和大学在校生、大中小学教师、专业教育研究人员分别单列,与以上分类相比,还对企业界经理职员、艺术家和非农个体劳动者进行适当归并。

表 7.4 不同界别写信者的来信频数情况

排序	写信者的身份	2000 年		2002 年		合 计	
		来信份数	占百分比	来信份数	占百分比	来信份数	占百分比
1	中学教师	36	23.5%	15	20.8%	51	22.7%
2	身份不清者	39	25.5%	12	16.7%	51	22.7%
3	经理等其他人员	20	13.1%	9	12.5%	29	12.9%
4	大学教师	16	10.5%	7	9.7%	23	10.2%
5	小学教师	7	4.6%	8	11.1%	15	6.7%
6	离、退休干部	9	5.9%	3	4.2%	12	5.3%
7	高中在校生	10	6.5%	1	1.4%	11	4.9%
8	专业技术人员	8	5.2%	2	2.8%	10	4.4%
9	大学在校生	3	2.0%	6	8.3%	9	4.0%
10	行政管理干部	5	3.3%	3	4.2%	8	3.6%
11	农民	0	0.0%	4	5.6%	4	1.8%
12	人大代表	0	0.0%	1	1.4%	1	0.4%
13	专业教育研究者	0	0.0%	1	1.4%	1	0.4%
	合 计	153	100.0%	72	100.0%	225	100.0%

注:表中的专业技术人员包括工程师、研究人员、编辑等人员,专业教育科研人员另计;经理等其他人员为企业经理、公司职员、艺术家、非农个体劳动者等,以下各表亦同。

表 7.4 在一定意义上体现了不同界别写信者的参与程度,他们在 2000 年和 2002 年的参与程度有所不同。总体上看,与贯彻 1999 年第三次全国教育工作会议的形势相比,2002 年 10 月以前的公众

来信总量低于2000年的同期水平，但是写信者的界别范围有所扩展，人大代表、农民和专业教育研究人员在2002年进入了写信者行列，于是2000年的写信者（除小学教师和大学在校生外）在2002年的来信份数比例均有不同程度下降。在两年的写信者中，中学教师来信的比例最大，占总数22.7%，从信笺的内容观察，这些中学教师大多来自农村、乡镇和中小城市的普通中学。大中小学教师合计的来信份数占总数39.6%，已经成为写信者的基本群体。经理等其他人员也较为积极，占12.9%，在校生和干部所占的比例不够高。身份不清者虽然占比例很高，但是，无法判断他们来自以上哪个阶层，也没有人自我表明来自商业服务人员、产业工人阶层还是城乡无业、失业和半失业人员阶层，至少说明这几类人员在通过寄信反映教育政策问题方面的参与程度是很低的。

表7.5则显示不同界别写信者在信中反映教育政策主题的多寡程度。中学教师和身份不清者仍然居于前列，当然，与他们来信份数比例高有直接关系。但是，除人大代表、农民和专业教育研究人员等信件数量偏少不计外，其他身份的写信者从表7.4到表7.5的位次会有变动，凡位次上升者，意味着写信者的信件份数少，反映教育政策主题的种类多（如专业技术人员从第8跃至第5，行政管理干部从第10升到第7），反之亦然（如离、退休干部从第6降到第8，大学在校生从第9降到第12）。

信访的主题话语变化特点

不同年份公众来信的主题词呈现出较好的一致性，但信件话语的主题趋于发散且有一定漂移现象。

由于各种条件的制约，各个社会阶层或者界别的写信者不可能都有机会以著述立说的方式占有公开出版物文本来阐述他们对教育政策问题的看法。为了表达自身关于教育政策的愿望与要求，他们可能选择的一个较为经济和可靠的途径，就是直接向国家级政府部门尤其是主要负责人写信。样本信件反映教育政策的主题词，可被归纳成为37个类别（表7.6），初步分析结果是：

表 7.5 不同界别写信者在信函中反映教育政策主题的频数情况

排序	写信者的身份	2000 年		2002 年		合 计	
		主题频数	占百分比	主题频数	占百分比	主题频数	占百分比
1	中学教师	59	29.1%	18	16.7%	77	24.8%
2	身份不清者	24	11.8%	20	18.5%	44	14.1%
3	大学教师	32	15.8%	6	5.6%	38	12.2%
4	经理等其他人员	28	13.8%	5	4.6%	33	10.6%
5	专业技术人员	13	6.4%	10	9.3%	23	7.4%
6	小学教师	10	4.9%	9	8.3%	19	6.1%
7	行政管理干部	6	3.0%	12	11.1%	18	5.8%
8	离、退休干部	13	6.4%	4	3.7%	17	5.5%
9	高中在校生	12	5.9%	3	2.8%	15	4.8%
10	农民	0	0.0%	10	9.3%	10	3.2%
11	专业教科研究者	0	0.0%	8	7.4%	8	2.6%
12	大学在校生	6	3.0%	1	0.9%	7	2.3%
13	人大代表	0	0.0%	2	1.9%	2	0.6%
	合 计	203	100.0%	108	100.0%	311	100.0%

表 7.6 样本来信反映教育政策不同类别主题的频数及其比例

序号	主题类别	2000 年样本来信主题			2002 年样本来信主题			合 计		
		频数	占比例	频数排序	频数	占比例	频数排序	频数	占比例	频数排序
	合 计	203	100.0%		108	100.0%		311	100.0%	
1	宏观教育改革	14	6.9%	3	16	14.8%	1	30	9.6%	1
2	教育体制改革	16	7.9%	2	8	7.4%	3	24	7.7%	2
3	教学改革	14	6.9%	3	10	9.3%	2	24	7.7%	2
4	素质教育	18	8.9%	1	3	2.8%	8	21	6.8%	3
5	学生减负	14	6.9%	3	5	4.6%	6	19	6.1%	4
6	高考制度改革	13	6.4%	4	3	2.8%	8	16	5.1%	5
7	校内人事改革	9	4.4%	5	7	6.5%	4	16	5.1%	5
8	考试制度	14	6.9%	3	1	0.9%	10	15	4.8%	6

续表

序号	主题类别	2000年样本来信主题			2002年样本来信主题			合　计		
		频数	占比例	频数排序	频数	占比例	频数排序	频数	占比例	频数排序
9	教材改革	14	6.9%	3	0		无	14	4.5%	7
10	高教改革	6	3.0%	7	6	5.6%	5	12	3.9%	8
11	教育方针	7	3.4%	6	4	3.7%	7	11	3.5%	9
12	学制改革	6	3.0%	7	5	4.6%	6	11	3.5%	9
13	重视德育	7	3.4%	6	3	2.8%	8	10	3.2%	10
14	儿童教育	6	3.0%	7	3	2.8%	8	9	2.9%	11
15	办学筹资	5	2.5%	8	2	1.9%	9	7	2.3%	12
16	教育战略	4	2.0%	9	3	2.8%	8	7	2.3%	12
17	人才培养	4	2.0%	9	3	2.8%	8	7	2.3%	12
18	乱收费	5	2.5%	8	1	0.9%	10	6	1.9%	13
19	教师素质提高	5	2.5%	8	1	0.9%	10	6	1.9%	13
20	创新思维培养	3	1.5%	10	3	2.8%	8	6	1.9%	13
21	学科建设	4	2.0%	9	1	0.9%	10	5	1.6%	14
22	布局调整	4	2.0%	9	1	0.9%	10	5	1.6%	14
23	拖欠教师工资	1	0.5%	12	4	3.7%	7	5	1.6%	14
24	民办学校管理	0		无	4	3.7%	7	4	1.3%	15
25	重视教育	2	1.0%	11	1	0.9%	10	3	1.0%	16
26	农村小学危房	0		无	3	2.8%	8	3	1.0%	16
27	人才引进	2	1.0%	11	0		无	2	0.6%	17
28	教育市场化	2	1.0%	11	0		无	2	0.6%	17
29	学术腐败	1	0.5%	12	1	0.9%	10	2	0.6%	17
30	教育投资效益	0		无	2	1.9%	9	2	0.6%	17
31	科技转让	1	0.5%	12	0		无	1	0.3%	18
32	择校问题	1	0.5%	12	0		无	1	0.3%	18
33	产学结合	1	0.5%	12	0		无	1	0.3%	18
34	基础教育	0		无	1	0.9%	10	1	0.3%	18
35	教师培训	0		无	1	0.9%	10	1	0.3%	18
36	大学生就业	0		无	1	0.9%	10	1	0.3%	18
37	农村教师缺编	0		无	1	0.9%	10	1	0.3%	18

我们对表7.6的2000年和2002年样本来信37组主题词频次分布数据进行了相关分析,相关系数为0.5355,通过了1%的显著性检验,表明两年公众来信在主题词方面保持着较高的一致性。若剔除某一年份为零的13组数据,则24组主题词的两年相关系数为0.5168,刚好通过1%检验,正相关程度要比37组弱一些。可以推测,在两个年份之间,公众来信近2/3的主要内容处于相对稳定的状态。

在两年样本来信涉及的37项主题词中,位于前10名的共有13项主题词,按照总频数排序分别是:宏观教育改革、教育体制改革、教学改革、素质教育、学生减负、高考制度改革、校内人事改革、考试制度、教材改革、高教改革、教育方针、学制改革、重视德育。这13项主题词的频数和为223次,占样本来信反映教育政策主题词总频数的71.7%。我们还采用模糊聚类方法对两年来信主题词频数进行聚合分析,结果表明公众来信的主题分布基本上围绕着相对稳定的主题展开,叙述的主题词范围构成一个比较稳定的话语谱系,主要集中在基础教育领域,尤以中学教育改革(诸如,学校体制改革、学生减负等)为主要信息点,少数涉及高等教育和其他教育领域,这一方面反映了当前在基础教育特别是中等教育阶段存在着大量热点难点问题的现状,另一方面也与写信者的身份构成有较大的关系。比较同期的《政报》文件主题可以发现信访提出的问题,比如关于"学生减负"的问题,在1999年教育部已经发布了专门的文件通知各地予以明令纠正和制止,但是在2000年仍然有为数较多的上访信件(频数比例占到6.9%)在反映这个问题,这可以从社会最基层反映出群众来信是对教育政策执行情况另一种信息反馈形式。

从2000年到2002年,样本来信的主题词频数呈现出逐渐发散的趋势,尽管在类别数量上仅从30项增至31项,但是,2000年按照类别分布的主题词频数的离散系数①为0.7798,可见该年的数据分布已很不均衡,到2002年更扩大为0.9341,接近数据差异非常显著的临界点。与此相关,两年合计的主题词频数的离散系数高达0.9145,说明不同背景的写信者所反映的政策主题相当分散。而且,即使对于相对集中的前13项主题词,2000年的频数和占总频数的74.9%,到2002年这一比例下降为65.7%。比较稳定的教育政策主题词分别有:儿童教育、办学筹资、教育战略、人才培养、乱收

---

① 离散系数=标准差/均值,此值越大,表示数据分布越分散。如大于0.15,表示数据之间具有统计意义的差异;大于0.5,表示差异显著;大于1,表示差异非常显著。

费、教师素质提高、创新思维培养、学科建设、布局调整、拖欠教师工资等,居于两年主题词频数合计的第11~14位,占总频数的比例一般在1.5%~3%。而重视教育、学术腐败等每年都有人提,但是仅占1%以下。信件反映的主题发散与写信群众的社会身份、地域环境、所面临的教育政策问题性质的差异有较大的关系。教育政策的问题话语是在特定的问题情境中孕育和产生的,其中也包括了写信人向政府诉求时所隐含着的各式各样的内心动机。这些动机是否代表民情民意是难以通过信件内容来鉴别清楚的,因此,对于信件内容的确认需要辅之以客观真实的实地调查验证。

样本信件的话语主题出现漂移,即使是一直列前10位的主题词频数比例,也随着社会基层教育热点问题的消长而有所变动。其中,宏观教育改革、教学改革、校内人事改革、高教改革、教育方针、学制改革等比例有不同程度的增大,显示了写信者关注宏观教育政策兴趣渐浓的现象,而其余项主题词比例均在减小,曾在2000年居于第7位的教材改革政策问题,到2002年已无人提及,或许这已经不再是信访反映的主要问题,因为新的问题在不断地取代老的问题。这在一定程度上也说明在全国教育工作会议精神指导下,近年来基础教育教材改革的进展较为顺利,虽然还有许多深层次问题需要在理论与实践中加以逐步解决,一些被政府关注并开始着手解决的问题,已从2002年的公众来信话语谱系中淡出了。

影响来信主题词的因素是异常复杂的。比如,2000年位于第一的素质教育下降到2002年的第8位,主题词频数比例从8.9%下降到2.8%,这种状况恐怕与教材改革不同,从素质教育的发展情况看,显然,在素质教育的实施方面,在全国范围远未收到预期的成效。按照2000年的主题词频数排序,与教材改革同时成为当年话题的人才引进、教育市场化、科技转让、择校问题、产学结合,在2002年的频数为零;而2000年尚未涉及的民办学校管理、农村小学危房、教育投资效益、基础教育、教师培训、大学生就业、农村教师缺编,在2002年开始有人关注,尽管频数占比例均不很大。由于写信者的基本群体是各级各类学校教师,所以他们对教育政策问题的叙述大多都采用了比较感性的材料,主要关注的是基层学校的或与公众利益关系密切的问题。当然,这些来信话语主题的显著漂移现象,还与地区的背景环境有一定关系,还须与同期的政府文件、专业学者讨论及新闻媒体报道等话语主题进行比较分析,才能判断其伴生性、同步性、超前性及滞后性的程度。

能够贴近教育政策话语的写信群体是接受教育管理知识程度较多的人

大学生、现职干部和专业教科研究者与同类写信者相比,其信件中所反映的提议更为贴近教育政策热点问题。从其身份特征看,他们是接受教育管理知识程度较高的社会群体。

对于按总频数比例排序的前10位13项主题词,不同界别写信者的态度也有所不同。在2000年涉及上述13项主题词的提议频数为152次,2002年为71次,我们不仅统计了不同背景写信者围绕13项主题词的提议频数及其所占比例,而且统计了某类写信者在13项主题词中分项提议频数,据此测算了平均的重要性系数,即由此可以判断某类写信者关注的问题是否贴近公众关注的热点。需要注意的是,写信者提议频数所占比例可以在不同年份之间进行对比,比例越高者,表明该类写信者参与政策的积极性越高,而重要性系数仅能用于同一年份不同界别写信者群体之间的相对比较,表明写信者关注问题贴近关键性政策问题的程度,跨年份的数据对比无大意义。

表7.7 不同界别写信者在前10位主题词的提议频数和重要性系数

写信者身份	2000年			2002年			合计		
	提议频数	占比例	重要性系数	提议频数	占比例	重要性系数	提议频数	占比例	重要性系数
人大代表				2	0.028	0.081	2	0.009	0.038
中学教师	50	0.329	0.094	11	0.155	0.169	61	0.274	0.064
身份不清者	17	0.112	0.089	14	0.197	0.146	31	0.139	0.061
大学在校生	6	0.039	0.081				6	0.027	0.06
小学教师	7	0.046	0.084	6	0.085	0.142	13	0.058	0.06
农民				9	0.127	0.279	9	0.040	0.059
大学教师	24	0.158	0.078	5	0.070	0.148	29	0.130	0.058
行政管理干部	5	0.033	0.125	8	0.113	0.115	13	0.058	0.056
专业教科研究者				4	0.056	0.159	4	0.018	0.055
经理等其他人员	22	0.145	0.078	3	0.042	0.11	25	0.112	0.054
离、退休干部	10	0.066	0.069	3	0.042	0.289	13	0.058	0.05
专业技术人员	8	0.053	0.072	4	0.056	0.229	12	0.054	0.049
高中在校生	3	0.020	0.084	2	0.028	0.081	5	0.022	0.048
合　计	152	1.000		71	1.000		223	1.000	

表7.7显示,2000年与2002年合计,中学教师凭借占主题词总频数27.4%的优势,在提议重要性系数上稳居第一,但与随后的写信者群体的差距仅在千分点范围内;大学教师所表达的提议虽占主题词频数的13%(居第3位),但是重要性系数却退至第6位。大学生与高中生在提议频数比例上相差不大,但是重要性系数排序相差8位,可见大学生的提议比高中生更能靠近政策问题。如果说人大代表的提议频数占比例过小,很难判断其提议重要性系数很低的原因,但是,作为同期写信者的对比,农民和专业教科研究者的提议频数比例亦不高,但是贴近关键性教育政策问题的程度是较高的。离退休干部与现职行政管理干部的提议频数比例相同,但在重要性系数排序上,前者比后者落后3位。以上分析尚未涉及样本来信的质量分析,即使所有写信者都觉察到某类教育政策问题特别重要,但是在围绕这一问题开展调研或撰写信件的过程中,他们表现出来的研究思路、技术路径、分析功力和信息依据等方面特点,可能差别更为巨大。此外,我们曾试用聚类分析方法,看某一界别写信者在某类提议主题方面是否存在着一致性,结果没有通过统计检验,估计主要原因是此类公众来信与具体案件上访信存在显著区别,大多数来信均涉及4~5个专题,比较发散。

公众来信的话语中蕴含着自发诉求教育政策的机制

社会公众想要把教育政策问题反映上去,就必须遵循行政组织的逐级报告制度。尽管有时能够偶然地越级报告,但其大致还是沿着从民间到政府程序化的途径来进行。袁振国(1996:51)认为:"以政策问题的确认为例,问题的察觉和描述是通过自下而上的报告体系来完成。如果中间层次太多,在这种自下而上的报告链条中,经过描述的问题传输到最高决策系统时(由于信息传递中的衰减和损失),可能与原来发生的客观事实大相径庭。"行政组织内部所制定的共同知识规范就会像一个有关制度的"筛子",不断地将报告话语主题进行浓缩和过滤,按照行政报告程序逐级向政策决策层报告。由于教育政策的制定是国家政治组织的行为,政治组织的体系结构基本上仍然保持着科层组织结构的特质。这种职能法定、层级明确、责任清晰、授权有限的科层管理模式,在传递和报告群众中内含的情绪化叙述和感性色彩的民意信息描述时,科层之间的行政管理关系、行

政组织规范和"职业禁忌"会促使行政部门主管对"来信"中所报告的信息进行必要的审核,尤其是对这种带有自发随意色彩且话语比较"敏感"的信访事件可能进行必要的"再叙述"(也就是行政管理用语中所谓的"集中起来、专门汇报")。这种"再叙述"的过程就不可避免地要将信访事件的信息情节进行突出"重写"或"淡化"消减。在此过程中,报告者对信访事件的知识、观点和态度会有意识或无意识地写入有关"信访的总结汇报"当中。报告者对于信访信息的"再叙述",对于信访信息能否被教育政策决策者收听到,并酝酿演变进入教育政策议题起至关重要的传递作用。事实上,由于教育行政管理目前处在有关"政务公开"的监督机制和教育政策制定和执行的监督机制尚未健全和完善时期,不同地区教育行政体制并不总能保证从基层开始向上逐级报告的有效性。

公众自发的、现时发生的大量信件语料,有可能较快地反映基层教育问题,将有关教育政策社会效应的信息回馈给政府,并提出关注的话题和解决问题的提议。从社会话语互动关系上看,也可在一定意义上用来检验公众对国家教育决策的理解、接受以及参与的程度。尽管本研究没有对样本来信进行文本内部词汇频率统计分析①,但是,仅从教育政策主题词的统计结果也可以看出,公众来信所叙述的话题基本上覆盖了目前教育改革与发展的绝大多数领域,从一个侧面直接反映了不同社会阶层群体对教育政策问题的判断和评说。

为此,我们拟从社会语言学的"话语秩序(the order of discourse)"角度,对公众来信的话语特征作些尝试性分析。法国社会文化批评学派的先锋福柯在论述"话语秩序"时认为,在任何社会中,话语一旦产生,立刻就按照一定数量的程序而被控制、选择、组织和再分配。这些程序包括外部控制、内部规则和评估知识的规则。这些程序就表达了话语秩序。任何给定的文本,仅仅代表或表达着它作为其产品的那种社会关系(Foucault, 1977:113-138; Burchell, Gordon, and Miller, 1991)。而英国批判话语分析(critical discourse analysis)学派的倡导者费尔克拉夫(Fairclough)认为,所谓话语秩序就是某一社会或机构的话语实践的总体结构,话语层面的社会秩序,以及社会文化实践在话语上的历史特征(Fairilough, Norman, 1995)。本研究所说的话语概念,不仅指书面的和口头的话语的总和,而且指延伸到社会群体话语模式的集合及相关的权力(power)与参与意识

① 关于词频变化比较分析的方法,请参见:村田忠禧,2002(3)

表述特点。话语只有参与进入社会交往活动过程中，才能够体现出话语的语用价值。信访是一种民间自发的政治诉求，更是体现一种自发的社会交往意识。这并不意味着信访话语会独立于社会文化存在，话语与社会文化的关系始终是水乳交融的①。信访话语的处境也是如此。

公众来信的话语秩序性分析

在公众来信的话语秩序中，势必长期交融着主流话语的技术化与非主流话语技术化成分并存状态。

公众来信的话语秩序，在公开身份与匿名的条件下是不同的。尽管信息技术革命导致的信息权势及其载体的变化，如互联网以多样化的形态出现的教育政策建议，并不是本研究讨论的重点（托夫勒，1991；时殷弘，2001；莫谢拉，2002），但是，当写信者在正式署名向国家级政府部门直接致函时，就自然被纳入一种特定的话语秩序，实际上就是在社会阶层中自下而上地“垂直传递”话语诉求，而且从基层（个人）开始，经过越级报告方式，直达宏观教育决策层面，为此，公开身份的写信者不得不考虑使用的话语方式和可能产生的影响。相比之下，从前面的统计分析中也可以发现，身份不清的写信者（类似网络上的建议者）似乎在提出政策建议方面更为宽泛和活跃。

公众来信的话语秩序，在语篇结构上的表现是比较复杂的。随着信息交流程度的增强与教育水平的提高，人们对身边发生的教育政策问题越来越敏感。写信者在希望意见和建议能被政府所采纳方面付出不少努力，但是，来信大多反映的是对教育政策的自我理解而且是感性理解。至少从225封样本来信看，既没有看到基于较为系统的调研基础上的分析，也没有看到图片表格等表达格式，大量的论据和少数的数据也未注明来源出处。因此，大量非专业性话语与少量专业性话语交织在一起，而且在专业性话语涵义上还存在广泛而明显的分歧，基本上属于结构化与非结构化文本的混合体，这是样本来信话语秩序一个十分显著的特点。

相比之下，专家进行教育政策研究的话语秩序，是在一定的学术圈内建立的，话语信息主要是“水平传递”的。并且，学术话语秩序是由各种规则和标准支撑的。一个成熟的专业（学科）需要达到技术化的职业、专业组织、伦理法规、知识体系、外部效益、社会认可以及高

① 本节部分分析思路参考了：叶起昌，2002-11-27

度自治等一系列标准(赵康,2000),促使学术话语变得慎重、精确和规范,能够将感性成分控制在特定范围。人们只有具备相当的专业知识,了解和掌握学科的范式(paradigm)①,才能读懂专业期刊的文本②,理解学术话语秩序的内涵,客观上造成学术话语秩序只能在有限的、专门化的知识环境中建立和运行,而公众直截了当、感诉夹议的话语秩序特征是专业化学术文本所排斥的。然而,面对教育的问题,多数公众写信者并未意识到介入教育政策的学术话语秩序的难度,而是单凭自我的认识和感觉在教育政策上也应当拥有发言权(话语权势),因此,公众来信的话语秩序的基础就是不同社会阶层群体对教育的直观理解。既融入了不同群体的价值观,又受到地域文化的影响;不仅包含着教育发展的历史特征,而且还带有即时性,反映出对当时某类教育问题的"情感性"直白。

公众来信话语技术化的发展趋向

在信访行为表现上,存在着两类"谏言专家"、两种不同教育政策问题话语,表现出信访语用"专业化"发展与"决策信息直通渠道"并存。

从话语技术化的倾向看,公众来信的话语秩序是否有向话语技术化迁移的趋势呢?英国社会语言学者费尔克拉夫(2003:200-222)认为,话语技术化应具备以下条件:①专家"话语人员"的出现;②在"控制"话语实践方面的转变;③不受情境制约的话语技术的设计和规划;④在话语中有策略和动机模拟;⑤朝向话语实践标准化的压力。由此可见,话语技术化的程度表明语言发展的成熟程度和语言所表达的社会规则的丰富程度。并且,费尔克劳在《话语与社会变迁》一书中写道:"话语技术在有关语言和话语知识之间确立了一种紧密的联系。它们的目的是某些预期效果……它们通过有意识的设计而引起话语变化……话语技术涉及模仿,特别是出于人际沟通意义和话语实践之策略的和工具的目的而进行的模仿……话语技术化是与策略的话语指向延伸密切联系……话语技术化似乎正在延伸……"(2003:200-212)按照上述标准评价,群众自有群众的语言特色,他们的信访也是遵循社会政治技术规范的表现。教育学术期刊

① 库恩(T. Kuhn)认为,研究范式是由一整套概念和假定所组成的看待世界、解答问题的基本方式。

② 在对专业性论文和通俗化文章进行比较时,迈尔斯划分了科学的叙事和自然的叙事两种不同的文本方式,反映了公众和科学家对科学行为的不同理解以及媒体在其中扮演的角色(迈尔斯,1999)。

上所登载的论文，大多应当归类于制度文化正式的话语技术化文本范畴。而目前的样本来信则是民间别样的话语技术化的表现，不完全符合上述正式制度意识形态所规训的话语技术化的条件，而是表现出一种来自社会基层边缘群体的亚文化语用技术。其话语秩序中是主流话语技术化成分与非正规技术化成分混杂。这属于“各尊其道、各善其美”式的社会文化多样性表现。在话语技术化的市场中，政治意识制度也应当为民间的非主流话语技术化发展提供一席之地。如若全都要求群众来信符合“专家话语”的特质，对于教育政策问题的社会信息反馈并非有利。有关全社会话语技术化统一的案例，我们从中国现代社会变迁中，可以找到那种全民族完全采用一种单调“高声”狂吼着要进行“四海翻腾云水努，五洲震荡风雷急”式的“高度理性批判”话语时期。从唯物辩证史的发展观看，一个所谓千篇一律的话语社会往往就必然包藏着非同寻常的语言交往危机和社会现实危机。因此，从社会语言生态平衡的发展看，群众信访即便是作为一类非主流的民间话语技术，也应当受到主流政治意识形态的呵护。况且，这种非主流的亚文化话语技术还在一直积极地学习和模仿主流话语正规的语用技术，以寻求社会对其身份和话语权力的合法性予以同情和确认。例如，群众来信的高峰时间段总是落在“两会”召开期间，或者是每年的岁末，这显然是与群众自发诉求的内容有关，“两会”期间信访诉求话语主要在宏观教育改革和教育的法律解释等的政策领域，而岁末则就要将切身的经济利益和生活所遭遇到的“不公”损害，诉诸于“公堂”，请求中央政府“为民做主”。在一个崇尚民主与文明的现代社会中，政治文化的法制化和开放程度赋予公民以合法的知情权与申诉权，社会的亚文化群体也应该有被赋予积极参与教育政策诉求的话语权力和给予合理宣泄社会不公意识的机会。并且，这种话语权力在凝聚社会交往的能量过程中会受到主流文化意识的整合。2002 年专业教育科研人员在写信者中的比例略有上升，个别专家“话语人员”开始向政府写信反映教育问题。一旦写信者的话语技术化程度提高到一定阶段，就会出现与公众来信话语秩序相分离的倾向。其中主要的原因，一是写信者将逐渐为学术话语秩序所接纳，或被媒体关注，再想反映教育政策问题，完全可通过公众来信以外更为便捷的渠道；二是如果写信者有可能将社会身份上升为“著名专家学者”，这就意味着其话语的正式话语技术化程度增长到了可以进入主流意识形态进行对话的地位，直接向高层决策负责人提交的意见或建议，往往被批转研究或批复本人，实际上

建立了与高层领导的信息直通渠道。但是这种“上升”需要归置于社会法制的监督限制之下,因为如若这种“便捷的”信息传递的渠道为公众知晓,并视为教育政策决策合法化的信息模板,这就给普通公众来信的话语秩序转向提供了参照的目标。以自发诉求话语形式来参与教育政策决策的机制将会被修改成为“单向度的”信息传播机制,教育政策科学决策的信息互动交往功能将渐趋于“单向化”和“特权化”,教育政策兼顾社会各个不同阶层人群利益的文化全面性将受到等级阶层的权力抑制,政策传播将渐趋于享有某种特定利益的“信息灵通”人士之间的局域信息交往,有关寻取教育政策决策信息“寻租”式的“交易”将有可能在教育行政组织和政策“可能的受益者”群体之间滋生,如此将使得教育政策的正常传播途径和贯彻执行过程复杂化,制度化的政策决策机制会受到“有信息直通能力的利益群体”的影响。教育政策决策信息“特权化”现象将产生,公众对教育政策的知情权和反映权将受到“消减”。这极不利于我国政府目前所奉行的改革开放国策,并应当予以警惕。需要建立一种制度化的教育政策信息反映机制,以利于全面真实地反映社会各个阶层的人们对于教育政策的态度。所以就必须建立科学民主公平的教育政策传播机制进行政策信息公开化的修复。自 1989 年教育部每年出版的《教育部政报》就是一种教育政策信息公开的机制。很遗憾的是,目前社会中众多教育政策分析人员和关注教育政策的人们却很少有耐心去仔细查阅现时的《教育部政报》内容,结果导致在众多学术期刊发表的有关教育政策的学术论文,其主题与《教育部政报》反映的问题几乎是不同语境中的两种话题。这就不免要让人对此产生疑问,目前汗牛充栋的众多的所谓“中国教育政策的……评论”等书刊或文章,评论研究的究竟是现实中真正的中国教育政策问题,还是“学术观念中”的中国教育政策问题,还是对教育政策评论的评论?要从社会叙事的观点上看,群众来信似乎在很积极、经济(就那“几页稿纸”和那“几块钱的邮资”)且直截了当地就他们所遇到的教育政策实际问题向政府写信,“要讨个说法”,并且就有关的利益用信访话语急切地向政府告状叙说着,“希望中央领导同志重视和关注……问题,并尽快能给解决……。①”

况且,在初始阶段的教育改革热点问题,并不总能采用标准化话语来概括,还有网络论坛等其他信息传播渠道对于公众意见的分流,

---

① 取自群众来信(编号:2000120005)内容。

这都决定了公众来信话语秩序的技术化程度会有所提高,但不可能很快。在相当一段时期内,公众来信势必是多种多样的话语共存、主流的技术化与非主流的技术化、技术化与非技术化的话语成分混杂交融的局面。韩毓海(2001:119-120)认为,有些知识分子“习惯性”地“对表述的内容与形式、对真实的表述以及表述本身进行规范和压榨,更以‘常识’的名义迎合并灌输到公众中去,从而助长了整个社会的‘共识’。历史上,不只一个人指出过,这种虚假的‘共识’是一个社会的最‘不良习惯’,是社会陷入僵化和惰性的最佳前兆。”同时,有些政府官员在一定的行政意识文化制度框架内,出自于本位利益的考虑,也可能经常就教育政策在决策和执行中出现的问题进行合法的“文饰”,并借助于“学术研究”成果的“学术权威话语效应”来合理掩饰政策决策中的失误。相形之下,群众来信似乎不大关注这些所谓的“常识”、“习惯性”、“共识”或“文饰”等概念符号,他们的来信更为注意是对自我感知的问题信息所关联的生活事件的评论。由于目前国家正在制定法律积极进行有关政府政策信息的公示制度(比如设立“政府新闻发言人”制度),在依法确立的社会舆论监督机制正在逐步健全的进程中,政府和民间对于教育政策制定和执行过程的信息掌握程度是不均衡的,群众也就可能在这种不对称的信息条件下,对教育问题产生一些片面、偏激的认识和话语评论。如此,将围绕着教育政策问题,有关的社会话语汇成了一种“各尊所闻、各行所知”的争鸣局面。尽管如此,即使公众来信话语秩序的技术化程度提高了,仍然在反映民间诉求上具有不可替代的作用。因为,群众书信是宪法赋予公民有关教育政策知情权力的表现。从社会进步的全局来看,群众书信与学术话语秩序保持着一定距离,不仅是完全正常的,而且是客观存在的,这是一种来自民间的有关教育政策的自发诉求的声音,是联系整个教育政策社会话语谱系的一种不可缺少的中介话语信息。同时,信访本身是一种文明社会中人类情感的合法宣泄与释放,是社会基层人群由于生活与各种利益困扰所产生的对于社会政治系统的意识形态主流认识偏差所致,积蓄的不利“精神能量”的合法宣泄与释放。用“来信”投票总比用“脚”投票或用“行动什么的……去投票”等要更加体现出“政治文明的有序发展”,更有利于维护国家的安全,也更能体现社会文明的进步程度以及社会民主与法制的有序性与稳定性。在一定的社会运行程度上讲,社会公众自发的言论诉求机制和谏言机制就是维系国家政府教育行政体系健康运行的“安全阀”,教育政策的制定者和政策的执行者万万不可

抱着“例行公事、悉听尊便”式的“官本位”态度漠然视之!

建立和保持一个有政策决策权威的群众信访信息回馈机制非常有利于维护全社会的和谐发展和长治久安。目前,中国在处于社会转型时期,行政制度的“民本化”改革是保持国家稳定的关键。从社会系统整体心态稳定调适的角度来讲,群众信访这种诉求机制应当受到国家政府政治文化制度的小心呵护。决不能简单地斥责为是“一群刁民”在“告刁状”! 在现代国家政治制度中,民意民情的诉求机制是作为政府执政管理的一种重要监督机制存在的。国家政策的决策集团在进行有关的政策设计时,非常注意全面倾听和综合社会不同界别人群的态度。教育政策的决策科学化要求决策者与公众之间能够保持一种信息畅通的“对话”交流。决策者应当时刻注意倾听来自大众的声音,从而实现教育管理者和群众、大众和精英之间的对话互动和相互理解。建立起维护社会稳定和发展的相互信任的信息沟通渠道。这正如德国法兰克福学派思想家哈贝玛斯所言:“在对话理论所设想的国家中,人民主权不仅仅体现在自由公民的共同的议会中。他应当返回到讨论和协议的多个主体进行交往的循环中,只有在这种自主的形式当中,人民主权的交往性的流动性权力才可以把国家机器的行政管理权力约束在人民意志的范围之内。”信访是作为对于政府的工作以非主流监督和反馈的形态存在的,尽管在相当数量的群众来信中,存在着言不由衷、语言随意、感情色彩浓重的语用现象,但信访行为本身表现了群众对于国家教育政策的一种自发的关注,我们从信访的调查中看到,大量的来信群众都以积极乐观的心态表述了对政府进一步加大教育改革力度,提高教育质量的期望。这从客观上反映了群众对中国教育发展的希望与期待。

公众来信中的话语谱系观

对教育政策执行情况自发反馈为核心语义构成了来信的问题域,促使公众来信形成了从微观到宏观、大量问题话语与少量抽象话语交汇的连续收敛型话语谱系。

从公众来信的整体角度进行文本归纳,可以看到大量与公众利益休戚相关的实际问题话语和少量表现教育理想价值观的抽象话语的交汇,形成从教育系统微观层面到教育决策宏观层面的连续话语谱系,这一话语谱系反映的是公众来信中暗含的语言意识水平。公众来信的语篇构成大体上可分为三大部分,一是礼貌用语,二是问题诉求,三是问题的阐释。

在第一部分的礼貌用语中，写信人试图采用有亲和力的尊称，搭建起与读者之间跨越层次的社会沟通关系，这样的尊称对于介绍整个信件的内容是必不可少的前提，其他礼貌用语都在辅佐构建交际话语体系中所谓的“积极的面子”，而这种礼貌用语是不可能出现在学术论文的文本当中。“两会”代表的“谏言”提案，在格式上也不存在礼貌用语的寒暄，因为提案本身就是宪法所赋予公民的权力和参政议政职能。除此之外，在反映政策问题的其他形式上来信与提案有着几乎相似的书写风格，都属于是一种来自社会的有关的政治诉求表达。

在第二部分的问题诉求中，剖析来信叙述的问题范围被称之为“问题域”。它是揭示公众来信话语在整个社会政治语言谱系中定位的关键。公众来信应当归类于社会政治文化的范畴（罗斯金，2001：149），其影响力在于来信中提出的问题。通过对样本来信主要内容的语意分析可以看到，每一封来信的正文都会提出一个或一系列事关教育方面的问题，体现着自发问题话语的特点。公众来信的“问题域”，主要由基调话语和热点问题话语两方面话语构成。通过相关分析和聚类分析等方法，我们发现在两年中基本上保持一致的相关变化的问题，约占总量70%以上，构成了公众来信问题域中比较稳定的基调话语，大量涉及教育管理和改革运行遇到的实际问题。同时，公众所关注的热点话题话语，大多属于两年间频次变化较大的问题，与教育政策调整和热点问题状态有很大关系，这部分话语似乎是从基调话语中“蒸馏”出来的。例如，2000年的“学校乱收费”、“学生减负”等问题，2002年的“拖欠教师工资”、“创新思维培养”问题，带有对教育政策明显的即时追踪性质。然而，与此相关的专业教育科研人员撰写的学术文本，并非按照问题话语链接而成，即使从问题入手，也会很快进入一定的学术话语秩序当中展开议论。

在第三部分的问题阐释中，写信者往往通过采用大量感性的经验、经历和感受来对教育问题进行阐释。也有少数作者就某一问题做过有限的调查（基本属于非专业性质的），引用专业文献或者报刊观点作为佐证，但此类信件所占比例不到3%，信件的内在逻辑性不是很强。许多写信者期待着政府收听到民间有关教育政策问题的反映，能够对他们的呼声做出回应，有些写信者则希望为调整现行教育政策或者研制新的政策出谋划策，但是掌握的基础信息和分析方法又不是十分到位。而学术文本必须有文献索引、调研方法及相关数据分析结果作为论点的依据，必须注重内在的逻辑性。

公众来信总体上体现了一种"持续关注"基调话语的特点,以维系公众话语谱系的连续性。同时,又从较广泛的基调话语中蒸发出少量热点主题话语,表明公众来信话语谱系是收敛型的,即公众来信话语谱系发端于现实问题意识,在民间实践层面发散,随信访时间序列,终端被政府所收听到,并且部分可能进入政策问题范围,逐步收敛于决策层面。虽然,教育研究的学术语言谱系从所涉猎的知识上基本呈现出发散型的话语,然而仅就教育政策问题诉求而言,也可能会针对某些教育利益冲突的问题迅速收敛,秉笔直书教育政策的问题所在。因为专家学者凭借专业话语权势的影响力,有可能实现从研究者到决策者的直接对话,促成实际教育问题直接反映到决策层面。公众来信话语谱系与学术话语谱系的收敛特征的差异,不仅体现在公众来信与专家学者在知识领域和信息依据方面的差异,更主要的是,公众来信关注的是现实教育中与切身利益相关的感性知识领域,而专家学者更为关注的是能触及教育系统规律的理性知识领域(北京师大教育政策与法律研究所课题组,2003-03-22)。

结论:自发诉求机制(个体)与国家教育政策之间的调适关系

社会公众来信通过对于教育政策的自发诉求机制,所体现的是一种写信人(个体)与国家教育政策之间就受教育的权利进行的话语调适关系。

以上所言的公众来信话语特征都似乎在"不谋而合"地揭示一个共同的诉求主题,这就是来自社会最基层的人们对于教育政策的利益关注。引发这种关注行动的是社会公众当中存在的一种自发的诉求机制在起作用。来信的话语秩序性以及向"技术化"方向建构的趋向,表明这种机制不仅具有"自发性"而且也具有某种"自学习"的功能,话语谱系则表现了这种机制在行政制度体制中所形成的一种话语活动的"踪迹(trace)"。自发诉求机制所体现的是教育政策与社会公众接受教育权益之间的调适关系。这种调适关系反映了教育政策在制定过程中并没有完全搬用来信中所提出的公众自发诉求。

正如上述所言及的,来信话语在维护国家教育管理的正常运行中,发挥着一种"安全阀"的"问题信息提示"作用,教育政策的制定中,仅是将公众来信话语作为了解认识民意民情的一项政策调研途径。实际在制定教育政策时,形成政策议题的话语构成成分复杂,往

往汇集了影响教育发展的各种政治力量的利益诉求，最终所形成的政策文本是经历了一系列的政治协商谈判所达到的话语均衡文本。对于此处分析需要说明的是，限于公众来信样本数量偏少和可用的分析方法过于分散等因素，我们仅能以公众来信分析为主线，选择了话语秩序及技术化程度、语篇中的问题域结构和话语谱系等方面，来初步分辨公众来信的话语特征，并在此基础上分析与专家学者话语特征方面的差别。对照同期的《教育部政报》可以看出，由于公众来信话语谱系中的浓重感性色彩和随意性的语体，使其与政府严谨的政策文本距离甚远；同时，专家学者话语谱系中纵横交错的逻辑推证，与简明的政策文本风格也不合拍。因此，在常规条件下，政府不可能仅依据某方面话语谱系匆忙做出针对性决策。在教育决策的科学化和民主化进程中，政府在关注公众来信的问题诉求和专家学者的问题争论时，政策文本在这里更像是一个“均衡器”，将于教育政策问题认定和决策酝酿中，谨慎地选用和参考两方面的话语。具体的教育政策行为在国家教育行政部门与社会公众要求平等接受教育机会的关系中，扮演的是一种社会有关教育权力利益均衡调适的功用。

我国实行人民民主专政的国体，这从制度上决定了公众依法自发有序地采用信访等言论形式参与政治的合法性。群众来信也历来为党和政府所重视。中共十六大报告提出：“改革和完善决策机制。要完善深入了解民情、充分反映民意、广泛集中民智、切实珍惜民力的决策机制，推进决策科学化民主化”以及“建立社情民意反映制度”。有学者认为：“信访是公民参政议政的特殊通道，也是弱势群体维权的有效法宝；信访是民怨的释放通道、民情的反馈渠道、民声的传声筒、民意‘用脚投票’的表达机制。”（刘武俊，2003-04-08）作者通过对信访行为的总体研究认为，对公众来信话语特征的分析，有助于促进管理层对公众信访中暗含的深层次话语意识的重视，从而可进一步提高决策层对相关教育热点问题的洞察水平，选择更加合适的决策模式来应对挑战。在决策的过程中，政府对来自一般写信者和资深学术专家的观点权衡方式可能有所不同，但是，有三个共同特征需要政策决策者注意，一是应当对公众来信中发生频率高的、公众反映强烈的教育话题进行认真的甄别和研究，对这些话题能否进入政策议题也可组织专门的论证。二是应当进一步建立和健全政府信息公开制度，保障公民的知情权，促使公众掌握更为充分的信息，以提升来信话语参与政策决策借鉴的技术化水平（杜刚健，2000；斯蒂格

利茨,2003)。这是政府“建立科学民主的决策机制”、“依法行政”和“民主监督”的重要环节。三是加强对信访信息接受与分析解决的综合联系,建立群众来信信息回馈机制,以体现政府勤政为民的办事效率和诚意,树立政府的公正信誉形象。党的十六大以后中央确立的“执政为民”的政府管理原则,中共十六届三中全会提出的“树立全面、协调、可持续的发展观”要求,2003 年 12 月落幕的中央经济工作会议更令人瞩目地提出了“以人为本”这样极富人文主义精神的政策话语,表示着政府现代执政理念的新思维与新方略。中国政府的国家政策发展决策观念正在向人本主义复归,具体到教育政策的决策机制建设上,一定要将信访信息分析与政府决策的信息采纳机制真正切实地联系起来,摸准中国教育发展现实的真正脉动,建立适合中国教育发展国情的教育政策问题信访反映机制,将教育政策在执行中的问题及时有效地控制在政府公共管理的有效解决框架之内,这是保证公共教育政策能够更有效地调整和促进我国教育改革与发展的一项非常重要的制度建设工作。它将有力地提高政府应用政策调控教育发展的执政能力,推动中国教育走向新的和谐发展境界。

第 8 章[①]

学校政策分析：一些办学机构的教育理念分析

本研究中，作者专门使用内容分析的方法考察研究了遍布美国的 267 个教育机构——包括小学、初中、中学和中学后教育的学校，并且划分了 4 个小组逐一进行各校的使命陈述内容分析。对原始数据进行自然编码后，我们得到 10 个主题词；接着再通过对总体样本进行分析，计算出每一主题词在这些教育机构的使命陈述中出现的相对频数。研究结果表明，学校教育的目标随着学校层级的提高而发生变化，从小学阶段强调学生的情感发展变化为大学阶段主要强调学生的认知发展。另外，使命陈述在各级学校内部以及各级学校之间也存在着很大的差异，大多数学校的使命陈述都集中为 4 个主题词，各学校关注的主题词各有不同。最后，作者还提出了政策性建议，以便于进一步的深入研究。

导 言

美国社会中，学校教育的目的是什么？不同级别教育的目标之间有什么差异？他们是如何相互强化的？就韦斯斯曼起诉波士顿拉丁文高中（Wessmann v. Boston Latin）和霍普伍德起诉德克萨斯州（Hopwood v. Texas）这两个案例而言，已经激起了人们关于学校教

① 本文专为提交给新英格兰教育研究组织（New England Educational Research Organization，NEERO）教育政策专题年会的研究报告。两位作者史蒂弗·斯蒂姆勒（Steve Stemler）和戴米安·贝贝尔（Damian Bebell）均为耶鲁大学教育社会与心理研究中心的博士。本文使用已获得作者授权。

育中高中层级和中学后教育的目标的争论。这两大案例都认为将种族作为入学标准之一的政策与完全根据考试分数决定是否录取的政策是相矛盾的。其实,这两项政策反映出人们因为看待问题的视角不同而对学校目标产生了不同的看法。我们可以分析各学校的使命陈述,构建出分析框架,继而理解全美教育机构所强调的不同主题。

研究开始之前,我们先提出研究的假设前提,并进行分析。我们的假设是:如果询问"学校教育的目的是什么",那么答案只能有一个,也就是说,学校教育只有一个目标。教育组织在继续讨论"教育的主要目的是什么"之前,应当了解众多教育机构现有的使命陈述的多样性和涵盖范围,这对讨论是大有裨益的。因此,本研究的基本目的是,确定各教育机构在使命陈述中表达出来的教育目的的总体范围。

因为哲学在关于美国学校教育目的的理论基础中地位相当重要,所以,我们有必要先概述这些基础知识。哲学领域中,很早就围绕着学校教育的目的展开了争论,本研究只想对之进行简单的回顾,目的在于为读者提供一个语境,帮助读者判断相关研究的进展状况,以理解并阐释研究结果。

哲学视角

一般来说,人们都期望学校能够根据自身的目标陈述来行动。泰勒(Tyler)曾说过:"如果教职工从组织成员的角度、从学校的规章制度和管理实践的角度,把所有的事情都看作是理所当然的,他们就非常可能参与这些事情,并建立一个统一协调的学校环境,使之有助于反映社会的态度。"艾拉希恩(Airasian)同意这个观点,他认为:"……任何关于教育目的的特殊观点,都足以导致教育界以及评价领域成为独特的组织系统。""当然,最终的分析结果是,教育目标是人们选择的结果,所以,它们必须是人们深思熟虑了的、适应于学校的价值判断。但是,在做出判断的过程中,人们必须综合地考虑教育哲学问题。"时间流逝,类似的哲学观点综合起来有如下几种:

坦纳和特纳(Tanner & Tenner)指出,"19世纪三四十年代,道德教育渗透进入整个课程教学体系。有关道德的描述为教科书添加了丰富多彩的内容章节,教师在教学工作中也开始注重学生的道德培养和性格养成。"玛瑞艾塔·约翰逊(Marrietta Johnson)则认

为:“学校总是宣称它的主要目的在于培养学生美好的道德情感和良好的性格特征。但是,抛开这些宣传不谈,我们注意到学校依然注重的是学生的智力发展,很多拥有美好的道德情感的人却无法毕业。”就是说,最古老也是最经常争论的哲学话题是“学校的目的在于促进学生感情的发展”。

然而,到了19世纪80年代,学校的焦点发生了戏剧性的转变,学校领域的主导理念成了:“保守派们说,教育的目的不是学习研究的方法,而是用人类记载的智慧去占据学生的大脑和思想。”具体地说,就是自由主义和个人主义的思潮对美国社会中的学校愿景产生了意义深远的影响。“个人的信念和获得成功的力量过去曾经是(现在仍然是)美国人熟悉的教义。自由主义思潮隐含的思想是,任何东西都不能限制或者阻碍个体的自由行动;成功,还是失败,取决于个人的努力……毫无疑问,应当通过修改课程以帮助学生从不适应转为适应,而这样做只能是使学校系统的目标落空而已,这个目标指的是清除那些不能够适应课程——课程的目标在于发展智力——的人。”因此,19世纪80年代的第二个哲学观点是“学校的目的在于促进学生的认知发展”。

“1909年,历史学家埃尔伍德·P·丘伯利(Ellwood P. Cubberley)指出,早在20世纪以前,学校便经历了一场目标和方向的变革。丘伯利说,‘学校一次又一次地承担起一个任务,就是:培养所有学生的社会感和政治感,统一学生的各种思维,统一学生的行为,最终为的是保存和发展我们的民主社会。’”“杰斐逊(Jefferson)希望创建一些国立小学,将年轻一代培养成忠诚的公民。另外,许多早期理论家也希望创建一所国家大学,培养国家领导精英并使之合法化。”本杰明·富兰克林(Benjamin Franklin)认为,学校课程应当更加功利化、应当为学生将来的职业做准备。我们面临的另一种观点就是,“学校的目的在于促进公民的有效培养并为学生将来的职业打基础。”

到了20世纪30年代,我们发现又出现了一个新的、与公民理论相交错的主题。“20世纪30年代,《教育前沿》(*the Educational Frontier*)杂志开始对‘教育的哲学陈述’的含义进行讨论,‘教育的目的仍然是帮助人们了解其生存的环境,以便人们能够控制环境而不是被环境所控制;另外,促使人们在控制环境的同时,创造一个更加有利于代际生存的世界。’”(Tanner & Tenner,1990,218)一个新的理念展现在我们的面前:学校应当培养学生的社会责任感,这种

责任感是构成有效公民的一个要素,同时它也暗示了一个成熟社会的进步。

有效公民的理念也可以与社会自豪感紧密相联。伯利纳和比德尔(Berliner & Biddle)指出:"为了适应社会的进步,美国的学校也可以为广大社会提供服务,诸如,采取某些形式,进行公众娱乐、社会启蒙以及社会精神培养。"(1995:262)R·M·拉托尼森(R. M. Latanision)对于这种看法持有异议,他认为学校的目的"是使个体有可能和睦相处并为着共同的目标而工作。"因为,家庭背景不同的学生,其间的差异性也很大,所以,可以通过让学生们参与社会活动、学校管理和体育竞赛来强化他们的社会自豪感(Tanner & Tenner,1990)。一旦思想中存有了这个目标,"课程设置的差异性越大,学生的社会融合程度就越重要。"(Tanner & Tenner,1990,122)此时,更新的观点就是"学校的目的是促进社会的发展、培养社会的包容性"。

显而易见,美国社会中存在着很多种不同的教育基本目的观。随着时间的推移,这些主要观点也在发生着变化。我们接触到的争论不外乎是说,学校的目的在于促进学生的情感发展、认知发展、培养公民和为职业做准备、促进社会发展以及培养学生的社会自豪感。因为这些理念依存的哲学基础大不相同,所以一所学校不可能同时采纳所有这些观点。这恰恰是作者致力于本研究的关键所在。如果我们发现各所学校在考虑学校教育对美国社会发生的基本作用时存在着显著差异,那么在他们的使命陈述上会不会反映出这种差异?

通过分析学校的使命陈述、考察学校的基本目的,我们希望能够引发关于美国大学和K-12学校的教育任务的对话。为了达成这个目标,我们建构了一个经验型的理解和分析教育使命陈述的框架。

使命陈述的使用

目前,存在着许多与一所学校的使命陈述以及这些使命陈述所包含的意义(或者意义缺失)相关的争论(Flake,1997),也有人对此进行了初步研究,系统考查了学校使命陈述的特性。哈瓦莱克和伯利尔1997年的研究工作考查了全世界52所耶稣会大学的使命陈述或与之相关的文件。两位作者使用演绎和归纳的方法,对这些使命陈述进行编码。他们主要关注的是人们观念中的耶稣会大学

的特征，方法主要是对不同的使命陈述中所使用的词语和短句的出现频率进行客观测量。斯特罗贝尔（Strobel，1997）也使用内容分析的方法，分析了被联邦中部学院及学校联合会（the Middle States Association of Colleges and Schools）所认可的120所学院或者大学的使命陈述。该研究的方法已经不再是对各大学的使命陈述中所使用的词语和短句的出现频数进行简单的计算。这几项研究的内容都是中学后教育机构的使命陈述。

在先前进行的教育机构使命陈述的研究中，因为其研究对象是作为样本的教育机构拥有的文献，所以这些研究耗费的时间和金钱是大量的。互联网的出现及电子技术的提高，促使越来越多的学校将资料展现到网络上——这个趋势在未来是有增无减的——使得我们可以通过网络接触到大量的文献并采集数据。这种获得文献的方法也不是没有局限性的，但它的确为研究提供了相当大的方便：不需要打电话或写信，我们便可以从网络上迅速地得到相关文献；另一个好处是，研究者可以采取积极的行动来收集资料，而不是被动呆板地等着别人回复他们的书面或口头请求。

研究假设

本研究的设计目的是，对我们收集到的学校使命陈述中表达出来的丰富多彩的主题词进行评价，探讨这些主题词在四组教育机构——小学、初中、高中及高中后——的使命陈述中的出现频率的差异性。经历了教育哲学有关文献的梳理，作者希望能够通过研究使命陈述来发现各学校之间的差异，及其随着美国教育层级的不同而系统变化的趋势。

本研究的两个假设似乎乏味得很。第一个假设是，并不是所有的中学后教育机构都企图完成相同的任务。换句话说就是，不同学校的使命陈述中显现的主题词在类型和数量上都存在着显著的差异。第二个假设是，美国的小学、初中、高中以及大学在进行使命陈述时存在着系统差异。虽然这两个假设本身非常简单，但其结果所隐含的意义应当能够对美国正规学校教育实践产生多方面的影响。

方法论

操作性定义

为了将研究聚焦在学校目标上，我们认为有必要在操作层面上

先对"学校目的"进行限定。就本研究来说,研究者将"学校目的"定义为"学校的使命陈述中所表达出来的观点"。尽管有很多文献可以供我们的分析所用,但是为了保证研究的一致性,我们仅仅选择展示在学校网页上的、明确标示为"使命陈述"的文献作为研究对象。

总体和样本

理论上说,我们的研究范围应当包括全美国所有 K-12 学校(共 108 000 所)和所有的四年制大学(共 1 900 所)的使命陈述。然而,实际上,我们是从这些总体中选取一定数量的学校和大学组成样本来进行研究的。抽取样本之前,需要确定研究的总体。我们共确定了三项标准来测定总体的实际大小:第一项标准是这个总体应当能够反映全美国大部分学校的特征;第二个标准是选取的学校应当能够反映 50 个州的特点;第三个标准则是我们能够通过网络上的链接进入所选取的学校的主页。很明显,并不是全国所有的学校都在网络上建有主页,因此,我们所抽取的总体就是这样的学校:它们是美国的 K-12 学校和大学,它们与两个综合网站中的任何一个之间建立链接(这两个综合网站中,一个是为 K-12 学校建立的,另一个是为大学而建立的)。

格利森—萨克曼(the Gleason-Sackman)网站(网址为:http://www.gsn.org/hotlist/)恰好符合我们的标准。这个网站上所罗列的小学、初中、高中学校的相关网页是最多的,该网站同时还设有一个索引,表明了逐步与该网站建立链接关系的学校的总数量。基于此,我们随即确定了研究的样本总体范围(1997 年春天)。抽取大学样本(1998 年春天)时,我们发现了一个非商业性网站(网址为:http://www.clas.ufl.edu/CLAS/american-universities.html),该网站上按照字母顺序罗列了全美国的大约 1 500 所四年制大学,包括公立大学和私立大学。由此,我们研究中涉及的总体(由小学、初中、高中和高中后教育机构组成)全部在这两个网站中可以找到。

然后,我们随机从这四组总体中抽取研究所用的样本。抽样过程大致可以分为三个步骤:

- 一个州一个州地逐一处理,通过计算机进行"复制"、"粘贴",将各组中列出的学校名称复制,形成微软表格处理文件。
- 按照表格处理软件,随机为文件中列出的学校编号。

- 将学校按照号码从大到小的顺序排列，形成四个独立的随机数据样本：小学、初中、高中和大学。

使命陈述的比率

研究的样本一旦确立，接下来就必须收集研究所需要的数据了。小学组中，我们浏览了 254 所学校的网页，其中 59 所小学(23%)有使命陈述可以用作研究。初中组中，我们浏览了 254 所学校的网页，其中 49 所初中(19%)有使命陈述可以用作研究。高中组中，我们浏览了 267 所学校的网页，其中 59 所高中(22%)有使命陈述可以用作研究。高中后教育机构中，我们浏览了 187 所学校的网页，其中 100 所(53%)有使命陈述可以用作研究。我们在浏览学校网站的过程中，发现一个现象，很值得我们思考，就是在大学的网页上一般来说能够很容易地找到使命陈述，而在小学、初中和高中的网页上就不行了。据推测，随着网络技术的持续发展，学校的网页也会越来越发达，毫无疑问，使命陈述的出现比率也会越来越高。表 8.1 是四个组中学校使命陈述的出现比率。

表 8.1　每组中使命陈述的出现比率

组　别	浏览了网页的学校数	拥有使命陈述的学校数	使命陈述的出现比率
小学	254	59	23%
初中	254	49	19%
高中	267	59	22%
大学	187	100	53%
总计	962	267	28%

程　序

使用自然编码的技术，人们编制了正式的编码手册。关于自然编码程序，菲埃罗斯、古莱克和惠洛克(Fierros, Gulek & Wheelock)在他们的著作中做了详细的解释。研究中，我们进行编码的程序大致如下：

1. 从每组中随机抽取学校的使命陈述构成研究的样本。

2. 最初抽取 60 个使命陈述(每组 15 个)，进行编码。两个研究者各自独立确定出样本中的主题词。

3. 两位研究者就这些主题词进行协商，达成共识。从这些主题

词中,我们可以清楚看出应用编码手册的痕迹。人们一般认为编码手册应当是两维的,每一个维度上都应当包括10个主题词,每个主题词又各自包含数量不等的次级主题词。

4. 在新建立的编码手册中,通过计算编码者的内在一致性及编码者外部的一致性来确定编码者的一致性程度(即信度,下文详述)。

5. 接着,作者对247个使命陈述(总数为267)进行两两配对,独立编码。

计算编码者的一致性程度

计算本研究中编码手册信度的目的,不是为了判断编码者是否就已有的或者缺失的主题词做出了"正确的"结论,而是在于,确定当编码者将各主题词应用到不同的学校使命陈述中时,他们在多大的范围内达成了统一的意见(即,编码者信度)。在本研究中,我们通过分析下述数据,来确定编码者针对编码手册中出现的主题词而达成的一致性程度。

1. 主要研究者之间的内在一致性信度(n=2);

2. 独立参与者之间的内在一致性信度(n=4);

3. 连续8个星期里编码者的外在一致性信度。

计算编码者的内在一致性信度和外在一致性信度时,有很多种量化方法。为了计算某一份编码手册的信度,最简单的方法就是计算全部编码者同意与否的百分比。然而,本研究中无法使用这个方法,因为,计算二维编码(缺失主题词对现有主题词)中各主题词的比例存在有一定的潜在困难,这又是因为,缺失的主题词在数量上可能会大大超出现有的使命陈述中显现出来的主题词。此时,我们使用科恩(Cohen)的K系数法(Kappa法)来矫正因为简单地计算同意比率而可能导致的系数的膨胀。值得注意的是,K系数与传统的信度系数在表述上是不同的。"必须注意,K=0并不意味着决定是矛盾的,进而就没有意义;K=0可能意味着决定不再一致,因为决定依据的是统计学上的独立数据。K系数值为0.2意味着我们可以观察到20%以上的一致性,并依此做出决定(Croker & Algina,1986,201)。克瓦尔赛斯(Kvalseth,1991)认为K系数等于0.61时,就可以"有理由"认为是全部检验通过了。

研究者们期望在研究群体外部找到一些人来进一步确定编码手册的效度。当研究者们公认的某些观点看法可能导致两个研究

者之间判定的编码手册的信度相当高；而外部人员对之评判却没有什么信度时，都应当这样做(Krippendorff, 1980)。

我们从每组学校中随机选取5个使命陈述来计算研究者的内在一致性。主要研究者(n=2)各自独立地使用已有的编码手册对使命陈述进行编码，比较编码的结果发现，他们在主要编码上的内在一致性系数(Kappa)总体上等于0.76(n=200)，其中，研究者对于小学、初中、高中和大学的使命陈述进行编码的内在一致性系数分别是0.72,0.91,0.55和0.86。

同样的，我们依然从每组学校中随机选取5个使命陈述，来计算独立参与者之间的一致性系数。我们将详细的书面说明分发给四个独立参与者，请求他们各自独立地使用编码手册上的主要词语对各学校的使命陈述进行编码。当我们对这四个独立参与者的编码结果进行比较时发现，独立参与者对所有主题词进行编码的一致性系数为0.65(n=200)，其中，小学、初中、高中及大学的一致性系数分别是0.69,0.77,0.49和0.63。

最后，我们选取了5所小学和5所中学的使命陈述来计算编码者的外在一致性系数(前后相隔大约8周)。对于前后两次编码的结果进行对照比较发现，编码者的外在一致性系数总体上等于0.80(n=100)，小学和初中分别为0.68和0.91。表8.2显示了编码的不同信度系数。

表8.2　对不同组别的命题陈述进行编码的K系数

信度类型	小学	初中	高中	大学
内在一致性信度 (2名研究者之间)	0.72 (n=50)	0.91 (n=50)	0.55 (n=50)	0.86 (n=50)
内在一致性信度 (4位独立的参与者之间)	0.69 (n=50)	0.77 (n=50)	0.49 (n=50)	0.63 (n=50)
外在信度	0.68 (n=50)	0.91 (n=50)		

因此，我们认为本研究对主题词编码的信度足够高，足以用来对学校的使命陈述进行有意义的、高质量的分析。

编码主题

表8.3是本研究中开发出来的编码主题。研究者使用自然编码和内容分析的技术，归纳出10个重大主题，这些主题均可以从各学校的使命陈述中找到对应的词语。请注意每一个词语都代表了

一种可测量的结构。

表 8.3 使命陈述主题编码

A 认知的或学术的 0—不确定(misc) 1—促进认知的发展或培养学术技能 2—培养解决问题的能力 3—发展创造力或者促进创造力的发展 4—发展研究能力	F 学校环境 0—不确定 1—自身环境 2—环境的一致性 3—以人为中心的环境 4—技术环境
B 社会的 0—不确定 1—促进社会交往能力的发展	G 精神领域 0—不确定 1—提供宗教教育和进行宗教教育的场所
C 公民的或职业的 0—不确定或者职业的 1—培养有创造力的公民 2—培养有责任心的公民 3—提供公共服务	H 当地社区 0—不确定 1—促进社区的发展 2—发展社区伙伴关系
D 身体的 0—不确定 1—促进身体的发展	I 全球联合 0—不确定 1—对多样性或者文化做出评价 2—增强全球意识 3—促进学生适应多样化的社会
E 态度的或价值观的或情感的发展 0—不确定 1—发展学生的积极态度 2—发展学生的伦理意识或道德感 3—促进学生乐于学习 4—促进学生终身学习 5—培养自我满足的学生 6—培养学生的自我约束能力 7—促进学生潜能的发展或发现学生的成功能力 8—发展学生的情感技巧 9—发展学生的自信心或自尊或自我价值观 10—促进学生的精神发展	J 教师或职员 0—不确定 1—环境的挑战 2—环境教育 3—提供有吸引力的工作或培养积极的学习者

结 论

表8.4列出了研究者对各组学校的使命陈述进行内容分析所得到的结果。通过自然编码以及测试了信度之后构建的分析框架，研究者对267个使命陈述进行了系统编码，并独立分析了四个学校组，然后，对学校使命陈述中出现的编码主题在各分类类别上进行百分比率的比较。下表显示了十项编码主题在各级学校的使命陈述中出现的频率，其中，“总体比率”指的是每一项编码主题出现的频率，代表的是这一项编码主题在一个使命陈述中单独出现的次数。表8.4还列出了这十项编码主题的次级编码词出现频率。因为一所学校的使命陈述中可能包含多个主题词，每一编码主题的“总体比率”反映了使命陈述中出现一个或者两个以上上述主题的学校的数量。本案例中，“总体比率”指的是一个使命陈述中某一特定主题词出现的总体数目，而不是所有主题词的总数。表8.4在显示编码主题百分比的同时，还揭示了各所小学、初中、高中和大学的使命陈述中出现的次级编码词。

表8.5展示的是主题词在四组学校的使命陈述中各自出现的频率。该表按照从高到低的顺序，将各主题词的频率排列出来。请注意，通过抽样分析可以看出，对于K-12学校来说，情感发展是一个非常重要的学校目的。还要请读者关注的是，随着学生的不断发展，学校教育的焦点是如何变化的：学校教育的层级越高，学校的使命陈述就越倾向于重视学术发展以及培养良好的公民，与此相对应的是，低层级学校的注意力大都集中在学生的情感发展上。

表 8.4 学校层面的主题

	小学	初中	高中	大学
	n = 59	n = 49	n = 59	n = 100
	百分比	百分比	百分比	百分比
认知的或学术的 = A				
0 不确定	2	10	7	10
1 促进认知发展	66	67	32	70
2 问题解决能力	14	16	8	12
3 创造力	5	8	2	19
4 研究能力	0	2	0	27
总体比率	69	76	38	83
社会的 = B				
0 不确定	8	14	3	4
1 社会交往	15	18	3	15
总体比率	24	33	5	19
公民的或职业的 = C				
0 不确定	7	8	7	17
1 有创造力	10	14	17	16
2 有责任心	19	24	15	21
3 公共服务	0	0	2	35
总体比率	36	39	35	68
身体发展 = D				
0 不确定	2	6	2	0
1 身体发展	15	16	5	11
总体比率	17	22	7	11
态度或价值观或情感 = E				
0 不确定	20	27	12	16
1 发展态度	12	20	7	4
2 伦理道德	7	6	8	12
3 乐于学习	8	8	3	3
4 终身学习	31	18	18	25
5 自我满足	2	0	3	2
6 自我约束	3	2	3	0
7 潜能发展	27	20	17	9
8 情感技巧	2	2	0	0
9 培养信心	24	14	7	1
10 精神发展	8	0	5	27
总体比率	81	76	52	62

续表

	小学	初中	高中	大学
	n=59	n=49	n=59	n=100
	百分比	百分比	百分比	百分比
学校=F				
0 不确定	10	8	7	9
1 安全	12	35	10	3
2 一致性	3	8	0	0
3 以人为中心	2	4	0	9
4 技术环境	5	2	0	6
总体比率	24	47	15	24
精神领域=G				
0 不确定	0	0	0	1
1 宗教教育或宗教场所	10	0	8	31
总体比率	10	0	8	32
当地社区=H				
0 不确定	5	4	7	9
1 促进社区发展	7	8	7	18
2 社区伙伴关系	19	24	18	23
总体比率	27	35	32	43
全球联合=I				
0 不确定	2	4	0	8
1 评价多样性	17	8	7	36
2 全球意识	7	2	5	10
3 适应性强的学生	7	20	17	6
总体比率	31	33	27	47
教师或职员=J				
0 不确定	15	14	15	15
1 环境的挑战	5	18	8	6
2 环境教育	20	20	10	5
3 提供有吸引力的工作	12	8	7	8
4 教师的质量	9	6	3	12
总体比率	41	51	32	34

表 8.5 各主题词出现的频率

	小 学 (n = 100)	初 中 (n = 100)	高 中 (n = 100)	大 学 (n = 100)
频率高低	情感的 学术的 教师	情感的 学术的 教师	情感的 学术的 公民	学术的 公民 情感的
	公民 全球社会 当地社区 学校	学校 公民 当地社区 全球社会	当地社区 教师 全球社会 学校	全球社会 当地社区 教师 精神的
	社会的 身体的 精神的	社会的 身体的 精神的	精神的 身体的 社会的	学校 社会的 身体的

另外,如果综合起来考虑这些主题词出现频率的话,我们会发现相同层级的各学校之间也存在着很大的差异。比如说,在"大学"这一组中(n = 100),各校的使命陈述中出现的主题词下出现频率在总体上呈正态分布,平均数为 4.23,标准差是 1.7(见图8.1)。换句话说,很少有学校仅仅强调十项主题中的某一两项。

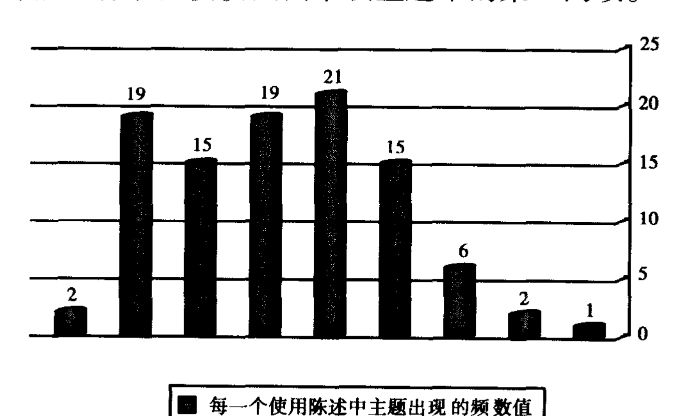


图 8.1 主题词频数分布图:大学的使命陈述中主题词频数分布图

为了更好地说明使命陈述的多样性,我们选取了四所大学的使命陈述作为样本呈现给读者。这四个使命陈述各有不同、各具特色,可以很好地显示出大学的使命陈述在内容上的丰富性。另外,

通过归纳这些使命陈述中出现的主题编码，读者可以更好地理解我们在研究中是怎样应用编码手册进行编码的。每个案例中，我们都对关键词进行了处理，标出其编码，用圆括号括起来。需要提请读者注意的是，有些句子没有编码，而有些句子却有好几个编码。这需要根据编码针对的文本来具体确定。

说明性案例

案例 1：阿帕拉契州立大学
(http://www.appstate.edu/www.does/depart/irp/factbook/asumssn.html)

使命陈述　阿帕拉契州立大学是一所公立综合性大学。大学开设一系列课程，可以授予学士学位、硕士学位，同时可以授予教育领导学专业的教育博士(Ed. D.)学位。阿帕拉契州立大学校园独具特色，教职员工(J0)具有较高的素质和丰富的专业技能。阿帕拉契州立大学将实践和进行学术(A1)传播作为自己的使命。为完成这些使命，大学特别强调教学(J0)、研究(A4)以及富有创造性(A3)和服务性的活动(C3)。阿帕拉契州立大学的责任是培养优秀的本科生和研究生，同时也为本州和本地区(H1)提供文化性服务和专业性服务。

编码：A1，A3，A4，C3，H1，J0

案例 1 是一个非常典型的大学使命陈述，共强调了四个主题词。分析阿帕拉契州立大学的使命陈述，读者应当理解为什么我们选择了特定的编码目录。主题词 A 代表的是认知因素，而其次级主题词 1、3 和 4 又各自强调了“促进认知发展”、“培养创造力”和“培养研究能力”；C3 项指的是“公民的或职业的”这一主题词，特别强调的是“公共服务”这一内在要素；H1 倾向于“促进社区发展”；而 J0 表示的是与“教职员工”状况有关的各种复杂情况。

案例 2 和案例 3，显示了不同学校的使命陈述存在着很大的差异。

案例2:日内瓦学院
(http://www.geneva.edu/graphics/whoweare/index/html)

使命陈述 日内瓦学院的教育目标是教育学生,为培养服务型领导(servant-leaders)、改造基督教王国(the Kingdom of Christ)(E10)提供后备力量。我们通过以《圣经》为基础的课程教学、优良的服务以及定期向基督进行宣誓(G1)等方式来实现我们的使命。我们的课程根植于大学的文科和理科教学,着重发展职业教育(C0),同时提供传统性课程和具有创新性的课程。

编码:E10,G1,C0

从案例2中,我们可以看出,日内瓦学院特别强调学校可以为学生提供宗教教育,它在强调"情感发展"这一主题时,特别关注的是"精神发展因素"。该校还试图传达的信息是,学生所处的团队是非常注重精神发展的。因此,即使学校鼓励学生精神和身体共同发展,学生也可以从精神团队中得到支持。尽管有所暗示,日内瓦学院并没有明确提及"促进学生认知发展",该学校的关注焦点主要是学生精神的发展。

与日内瓦学院不同,菲尼克斯大学的使命强调的是"认知或者学术",请看案例3。

案例3:菲尼克斯大学
(http://www.uophx.edu/uop/ourunive.htm#OUR MISSION)

我们的使命 菲尼克斯大学是一所非营利性私立高等教育机构,它的使命在于为已有工作的学生提供高品质教育。大学充分了解全社会的教育需求,通过包括远程教育技术手段在内的创新性教学方法,为全国各地的成年工作者提供接受教育的机会。菲尼克斯大学的教育包括一般教育和专业教育两种,可以帮助学生确立个人发展目标及专业发展目标。

菲尼克斯大学崇尚的教育哲学体现为参与、合作以及通过策略研究(A2)来解决问题,教职员工们进行了大量的学术知识储备并拥有丰富的专业实践经验,他们通过课程教学中的实际应用理论(J0),帮助学生整合各种学术理论、发现解决问题的各种策略。菲尼克斯大学既重视学校本身学术投入的有效性问题,也非常重视学生们的学术发展(A1),并将这二者的发展作为学校进一步提高学术和教学质量的动力。

编码:A1,A2,J0

有趣的是，菲尼克斯大学在其使命陈述中非常强调“远程教育”。这不奇怪，因为，“认知发展”在该使命中占据了统治地位，而且，该校在强调“促进认知发展”的同时，还强调“培养解决问题的技能技巧”。

最后，我们给出案例4的目的在于：揭示虽然一些大学在使命陈述中表现的是高度专业化倾向，但是，在极少数情况下（低于10%，见图9.1），大学的使命陈述中也会出现很多主题词。

案例4：约翰逊·C·史密斯大学
(http://www/jcsu.edu/purpose.html)

约翰逊·C·史密斯大学的使命是为来自于不同种族、不同社会经济阶层以及不同地域(I1)的天才学生和高动机学生提供高质量的教育。约翰逊·C·史密斯大学为学生提供自由发展的教育，同时也进行特殊领域的研究、为学生的进一步深造以及从事特殊职业工作做准备。大学努力培养能够有效地与人交流合作、具有批判性思考能力和独立学习能力、在自己的领域里独当一面的研究生(A1)。史密斯大学的校园环境(F0)，可以满足学生身体的(D1)、社会的(B1)、文化的、精神的(E10)以及其他方面的个人需求；这种环境能够培养学生的社会意识和公民责任感(C2)，有助于学生走向领导岗位，并服务于(C3)我们身处的这个生机勃勃的、具有多元文化的社会。因此，史密斯大学的责任就是向广大的社区(H1)人民提供领导、服务以及终身学习(E4)的教育。史密斯大学认为教学效果是衡量学校发展的首要指标，所以大学投入大量经费以招募和保持优秀的师资力量(J0)；为了做到这一点，史密斯大学努力促进教职员工的发展，鼓励教职员工参与研究及其他富于创造性的活动，并认可学术自由的原则。为了保持大学的整体性和稳定性、保存大学丰富的遗产并使之发扬光大，史密斯大学坚定地认为经费和人力资源是保障大学发展——大学发展的品质标志包括学生、教师、职员、管理者、学术及其他项目、设备、大学的运转以及校园环境——的基本要素，并保证经费和人力资源的来源。

编码：A1，B1，C2，C3，D1，E4，E10，F0，H1，I1，J0

简而言之，约翰逊·C·史密斯大学强调的是大学实践的各个方面。然而，有趣的是，十项编码主题中唯一一项没有被囊括的是“精神社团”这一主题词。该校的使命陈述提到了学生的精神需要，但是没有说明学校是一个强调精神发展的组织。

在阅读完这些使命陈述之后,问你自己几个问题:这些使命陈述和学校的教育教学实践是一致的吗?这些学校里的学生是什么样的?这些大家公认的主题要素在形成学校特色的过程中起了什么样的作用?如果这些学校要求申请者参加学习能力倾向测验(SAT)以获得入学资格的话,你会感到奇怪吗?

讨 论

人们普遍认为,很少能够在学校的使命陈述中发现关于学校的“真正”信息,人们经常将学校的使命陈述看作是花言巧语、不能表达任何意义的空洞言语。如果这是真的,那么对使命陈述进行研究,应当发现所有的使命陈述基本上说的是同一件事。本研究的结果与之完全相反:各所学校的使命陈述都是与众不同的,这些使命陈述和学校的系统结构一样,可以反应出各所学校的理念和运行方式。进一步说,通过使命陈述,我们可以窥视出这些学校在建校早期就创立了并公开宣传的教育目的。

本研究审查了我们选取的267所小学、初中、高中及大学的使命陈述。研究人员对抽样学校的使命陈述进行自然编码和内容分析,发现了一个基本事实:一般来说,学校在使命陈述中强调一个或多个主题词,这些主题词包括在研究人员编制的十个主题编码内。使用这些主题词,我们分析了各学校使命陈述的相近之处,同时也分析了它们不同的地方。本研究还分析了随着学校教育层级的提高,十项编码主题在各学校的使命陈述中出现频率发生变化的情况。

重大发现

第一个重大发现是,各组学校之间以及各组内部学校与学校之间,使命陈述中出现的主题词在频率以及类型上存在着很大的差异。由此,我们可以认为,假设所言的——“所有的学校在其使命陈述中都倾向于强调同样的事情”——不成立,虽然大多数学校在使命陈述中都趋于强调4~5个主题词①(大多数学校的使命陈述都着重聚焦于认知发展、情感发展以及为学生成为有效公民或者为从事职业工作做准备)。在每一组学校的使命陈述中,这十项主题

① 参见学校使命陈述的内容。

词出现的频率大致呈现为正态分布。

虽然所有的学校都有各自占主导地位的主题词，但是就主题词在使命陈述中出现的数量及综合性来说，每一所学校似乎又都有自己的独特之处。本研究发现，学校的教育目的各有千秋，各所学校都在强调其与众不同的、足以反应自身特定价值观体系的教育目的。因此，本研究的第一项假设——各学校的教育目的是不同的——通过这些经验性的数据得到了证实。

引人注目的问题是，人们经常挂在嘴边的宣传性词语，如“终身学习者”、“公共服务”、“社区伙伴关系”、“评价的多样化”以及“研究”等，实际上在学校的使命陈述中出现的频率不超过36%。这进一步说明了学校的使命陈述并不都是在说着同样的事情。同时，我们有必要思考的一个问题是：也许有些问题在学校使命陈述中没有明确说出，而它们与已经标明了的问题具有同等的重要性。我们还可以从研究中发现一个事实：除了“精神家园”以外，确实有一所学校提及了其他的所有主题词。

思考我们的第二项假设，不同层次的学校在使命陈述上存在着一定程度的系统性变化，分析结果证明这项假设是成立的（本研究的结果显示，正如各学校其在使命陈述中所表述的，学校教育的目的都在于促进学生的发展，不过各自关注的发展点却各不相同）。其中，81%的小学在其使命陈述中强调情感因素的发展，而只有62%的大学强调这一点；与此相对应，69%的小学在其使命陈述中强调认知因素的发展，而83%的大学强调与认知相关的因素的发展；中学的状况介乎于二者之间：76%的中学将“情感发展”作为主题词写在其使命陈述中，同样比率的中学认为“认知发展”是其使命陈述中的主题词。

更为有趣的是，我们来看看不同层级的学校在使命陈述中是怎样综合地使用这些主题词的（见图8.2）。结论是：学校的管理者并不是不假思考地把这些主题词写进学校的使命陈述中的。如果是这样的话，我们所选取的样本中应当全是这些主题词。但现在的事实却是，只有极少数主题词（n = 2）在使命陈述中的出现频率超过了50%（这两个主题词是“促进认知发展”和“教师”）。

从结果还可以看出，不同的学校组之间，次级主题词出现的频率也是不相同的。比如，“情感发展”这项主题词包括自我满足、自我约束、潜能发展等十个次级主题词，比较大学与其他学校的使命陈述：“法制意识”的出现频率是急剧增加的，而“行动潜能”和“内在激励

(培养信心)”这两项的出现频率,从小学到大学直线下降。从属于“情感发展”的各项次级主题词的频率变化状况如图8.2 所示。

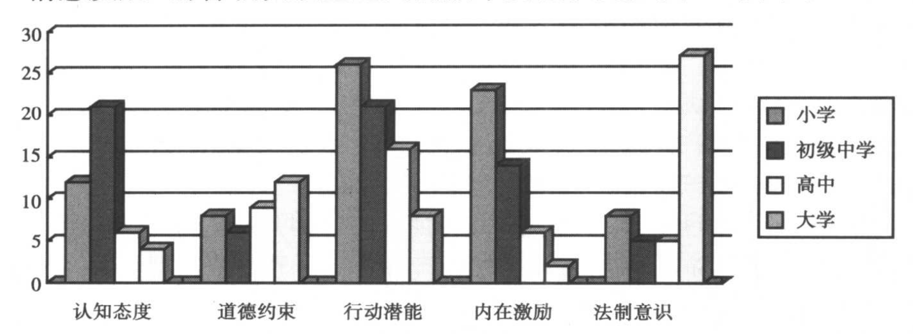


图 8.2 组际间“情感发展”主题词的变化情况(情感发展因素)

研究者也发现了其他各项次级主题词(公民关系、环境因素等)组际间变化情况,为了简便起见,此处不再论述。

局限性

行文至此,我们有必要讨论本研究的局限性问题了。第一个批评是针对作者对学校目的进行的操作性定义而提出来的。我们承认,仅仅对学校的使命陈述进行分析存在着将问题简单化的嫌疑,特别是我们还假定,所有的学校都把它们的使命陈述看作是对学校目的的反映。然而,我们认为,使用一个常用的测量标准进行研究是可以获取一些东西的,而且,使命陈述至少存在着表面效度。

本研究的第二个局限在于,没有分析学校对于某些主题词的关注程度。为研究而编制的编码手册本身包含多个次级主题词编码,但是编码手册没有对之进行具体分析,以致编码者无法辨别使命陈述中出现的某项次级主题词的重要程度。研究者们倾向于将本研究作为一个基础,目的在于进一步对教育使命陈述进行经验性文献收集和分析。如果我们继续研究的话,我们可能就会去探究各学校的使命陈述中出现的或者缺失的主题词的重要程度问题,因为本研究已经发现不同的学校之间确实存在着差异。

第三个批评针对的是:本研究在计算编码者的一致性程度(见表 8.2)时,只对十个主题词计算了编码者的内在一致性信度和外在信度。当我们收集到全部数据以后,才发现众多次级主题词也可以很好地帮助人们理解学校的使命陈述。在将来的研究中,应当针对次级主题词计算编码者的一致性信度,这对于最终的数据分析是大有裨益的。

其他的批评大都集中在样本的抽取上，因为我们抽取的样本全部来自于中小学或者大学网站上的使命陈述，那么，如果一所学校没有网站的话，它的使命陈述就不可能进入我们抽样的范围；而且，我们也只能分析那些我们找得到使命陈述的学校。从方法论上说，我们可以进行抽样研究的学校的比例就很低（见表8.1）。导致这个结果的原因有很多，例如，一些学校，虽然属于我们的研究总体，但是因为其网络链接不好用，或者服务器有问题，我们便无法浏览它们的网站；还有一些学校，网站上没有使命陈述，但是有一些相似的文件，诸如学校的任务、目的或者目标等，而这些又不属于我们研究的范围。因此，很有必要将通过互联网获取的使命陈述的学校比率同哈瓦莱克和伯利尔1997年的研究做一番对照，这两个人是通过信件来获取学校的使命陈述及相关文件的，他们的研究中该比率为38%。本研究中该比率值——小学23%，初中19%，高中22%，大学53%——与本领域中早期研究的结果相差不大，当然还应当考虑到本研究的标准比先前的研究更加严格——我们只分析各学校的使命陈述。

如果不考虑本研究中出现的上述种种缺憾，可以认为，我们运用计算机时代的研究方法论，为教育领域热点问题的分析提供了一种新的视角。网络可以成为研究者进行研究的新的工具，我们的研究也正是因为有效地利用了网络，才获得较大的收获。这说明，以网络为基础的研究似乎发展潜力巨大。

结　论

本研究的导言部分概括了有关学校目的的几种哲学观点。四大主题词——情感发展、认知发展、社会性发展以及为公民或者职业准备——曾经似乎统治了类似的研究。人们发现，随着时间的推移，这四大主题词中的任何一个都不可能成为学校的唯一目的。换句话说，美国学校和学校教育的目的对于不同的人来说就意味着不同的教育价值观取向，而且这种价值观取向可以随着时间的变化而变化。

本研究的结果表明，现在，在从小学到大学的不同层级的学校中，这些不同的观点仍然存在。研究结果还显示，我们可以从学校的使命陈述中归纳出占据重要地位的主题词。正如我们前面所论述的，当前，在K-12学校的教育中，最关键的主题词是情感发展，大学则在其使命陈述中认为智力发展是首要的目的，情感发展的地位就稍稍低一些。通过本研究，我们很容易看出，从K-12学校到大学，其

学校目的强调的中心是从“情感为主”向以“学术为主”偏移的。

有趣的是,教育界当前对于有关 K-12 学校的政策讨论,已经偏离了学生的情感发展,而转向认知和学术的发展。看看当前流行的学术规范行为运动、全面发展行为运动以及增加州的测验项目等活动,便很容易看到这个发展趋势。与此相类似,许多州现在都要求学生通过毕业考试以获得证书,这也顺应了这种变化潮流,即 K-12 学校强调的重心正在远离情感的发展,而走向认知和学术的发展。我们在学校的使命陈述中还没有发现这种变化迹象,如果将来再对 K-12 学校的使命陈述进行分析的话,可能就会更加清楚地看见这种变化了。

贡 献

使命陈述的有效性和内涵,以及使命陈述的实用性和简洁性,使得使命陈述成为运用经验方法分析学校目的的有用文献。如果说本研究做出了一点贡献的话,那么我们希望是:通过描述使命陈述的特性,我们“揭开了使命陈述的神秘面纱”;本研究还证明了,一旦编码者一致性程度或者说信度达到了可以接受的水平,我们完全可以通过自然编码、内容分析的方法对使命陈述进行分析。

本研究的结果引发了一系列有趣的政策问题。其中有些涉及大学的入学问题,例如,如果一所高等教育机构在使命陈述中描述了他们希望招收的学生类型,他们会使用什么样的标准来判断某个学生是否适合于其学校文化呢?如果使命陈述描述的是学校希望培养的学生类型,他们又将使用什么样的标准来衡量学生并保证学校完成使命?还有,直接将具有不同使命的学校进行对比,意味着什么?我们来看看《美国新闻及世界报道》(*the U. S. News and World Report*)上的排行榜划分:在不清楚这些大学努力达成的终极目标是否一致的情况下,对所有的大学进行排行,这种方法可取吗?本研究的结果显示,先对一所学校完成其既定使命的状况进行评价,然后比较那些使命相同的学校,这样的做法可能更富于逻辑实际,说服力更强些。

本研究结果告诉我们,对于学校使命陈述,确实可以不仅仅是从修辞学角度来研究,而且可以使用测量研究的方法。使用诸如内容分析之类的经验研究方法,我们可以观察到美国社会学校目的的变化状况,继而引发人们关于美国教育应当走向何处的讨论。我们的研究就是为这样的讨论提供了一个经验型的实证研究基础。

第 9 章①

21 世纪初期社会话语镜像中的中国国家教育政策调整机制

前　言

中国是世界上教育规模最大的国家,据国家教育部有关的统计公报,2001 年各级各类学校达到 1 350 000 所,在校学生已逾 3.2 亿人(这还没有将那些正游离于校外,在家庭或者在其他非正规教育场所接受形形色色"教育"的"学生"们统计在内)。概言之,中国大约平均每 4 个国民中就有 1 位是学生。从中国目前人口规模增长上讲,教育事业已经成为政府管理中继"三农"问题后影响中国社会发展的第二大要务。教育政策作为调整教育事业发展的国家行为和政府行动准则,涉及社会各阶层群体的利益诉求,由于教育政策的多重社会文化特征,以至于一项教育政策出台就有来自不同社会群体的多重话语评论涌现。社会有关教育政策的这种评论话语构成具有社会舆情镜像反映的社会民意特征。

教育事业从本质上讲是属于社会上层建筑中意识形态核心的文化事业,有关教育政策的研究就无法回避从社会公众的舆情话语反映来认识、看待、理解教育政策的内在规律。因为脱离社会话语实践,从单纯经济学和哲学角度的思辨分析,是难以直白地说清楚教育政策话语权究竟是面向人民大众的教育服务,还是面向少数经济富裕人群的"过渡教育"服务。而社会公众有关教育政策的评论话语,尽管是处在专家话语与民众言论、政府政策文本与政策评论、

① 作者为李钢教授。

积极的谏言与犀利的批评等相互夹杂着构成的社会话语谱系当中,然而这个社会话语谱系却从民意的反映途径,给现行教育政策的价值行为提供了一个社会话语的镜像,并反映教育政策执行过程中的社会认知情况。

从社会话语的活动来认识有关的教育政策发展机制,首先要确定一个分析范式。即如何从社会话语活动的角度来定义分析单元,也即定义一个社会话语谱系中的教育政策的公共话语空间。这个公共话语空间,要有别于一般公共政策的话语空间。众所周知,教育政策是政府制定和实施的非常重要的公共政策。由此教育政策也具备公共政策所有的通用价值观特征,比如公益性、公平性、权威性、强制性,等等。但是由于教育政策涉及人的发展与社会文化发展中知识传承的"意识规则"问题,这是一种社会内在深刻的政治文化"规训",因此教育政策活动又表现为如下特殊性:第一,教育政策的制定是一项成千上万人参加的极其复杂的工作,具有专门的术语,政策内容广阔多变,真正理解教育政策的本质并不容易;第二,相对而言,公众对教育系统比对其他社会系统更熟悉,对教育政策更加关心,大多数公民对教育系统都有一定的接触和了解,并且都积极参与教育活动和学校的管理;第三,教育领域的权力比其他任何领域的权力都分散,教育政策不仅在不同的权力级别上形成和执行,而且在同一权力级别上也存在复杂的权力分配,从而加大了教育政策活动的难度和复杂性;第四,教育过程本身具有多种目标,在不同的教育机构中又具有含糊性,评估教育成果具有先天的困难性,等等(那格尔,1990:442)。现实中国的教育政策话语评论的状况证实了弗莱德·S·库姆布斯(Fred S. Coombs)的上述归纳和判断。近年来有关教育观念的讨论,都往往与教育政策的调整相关,而教育政策的调整又往往伴随着多种社会话语评论回应,这种现象几乎成了每一个重大教育政策实施后所面临的共同境遇(李钢,2004:7),以至于曾主管教育的副总理李岚清(2003)同志在卸任后,著述评论新时期以来的国家教育政策变化时,感言道"人人都可以是教育评论家"。

一项教育政策竟然"触发"这么多种类的话语评论,究竟是教育政策的哪些文化特性招致了社会对教育发展的关注,有关教育政策过程表现为怎样的话语互动?教育政策为什么是这样表述而不是那样表述,教育政策的话语传播有什么特征?社会有关教育政策的评论话语与教育政策的制定有什么关联?从话语活动反映出哪

些决定教育政策发展的文化要素？这些归结起来就成为这样的问题：改革开放以来，国家教育政策遵循着怎样的发展方式，为什么会选择这样的方式？教育政策过程中含有哪些机制的作用，这些机制怎样影响着教育政策的制定？这些都是研究和理解教育政策应当考虑的问题，本研究拟从有关教育政策文献和相关的社会话语文本的分析入手，初步阐述国家教育政策过程的有关机制，以增强公众对于中国国家教育政策的理解和认识。

研究假设：教育政策的公共话语空间构成

确定一个教育政策的公共话语空间，这是本研究确定的从社会话语活动的角度理解国家教育政策的分析基础。所谓公共话语空间，就是有关在教育政策过程中的各类社会评论话语系统的构成，也就是有关教育政策的社会话语系统。回顾我国自1979年改革开放以来国家教育政策发展的道路，教育政策的话语主题从“教育革命”到“教育改革”到“教育改革和发展”到“坚持科学的发展观指导，促进教育的和谐发展”等，教育政策的决策思维范式也在相应地发生转型，从以政治话语作为教育政策的权力话语空间转型为政治与经济相结合的公共话语空间。教育政策决策的理性化和科学化程度在政策运行的实践中不断提高。按照西蒙的“有限决策理性”思想来定位，中国教育政策过程正处于一个“有限理性”的决策转型阶段。这个“有限理性”寓意着政府和全社会对教育事业的理解正在从群体局部利益的调整向整个民族和社会文明的价值观取向转变。林德布洛姆（林德布洛姆，1987：67）基于公共政策决策渐进理论，将公共政策过程划分为五个阶段：即情报收集、政策议题、政策制定、政策执行和政策评估。教育政策的公共话语空间在这五个阶段都具有影响政策决策的有代表性的话语构成。由于本研究所涉及的话语分析属于回溯性的后测分析，所收集的分析语料是内容相对稳态的文本，并且仅以官方公开发布的政策文本和可以通过公开合法途径查阅到的文献为依据。这样建立的教育政策话语系统应当包含两个部分：①教育政策的话语主体，即谁在“评论”和“言说”教育政策。这个主体集合（教育政策的叙述者群体）是由教育行政管理（决策者）者、关注和能参与教育决策的人群、研究教育政策的专业人群、传播教育信息的公共传媒人员、关注教育政策话题的社会群体等。②有关教育政策的文本，即教育政策的话语形

态,确切地说是一种文本形态。这是由上述人群对于教育政策所表达的有关话语文本体系所组成(这个系统如图9.1所示)。

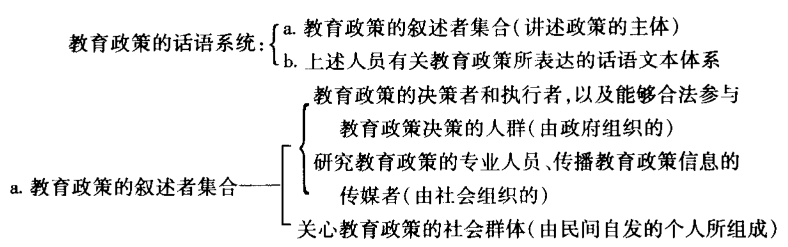


图9.1 教育政策表达者谱系图

在研究文本的选取原则上考虑到以上话语空间中的叙述主体几乎分布在社会的各个阶层,他们对教育政策的话语分散在社会文化传播的各个角落,作者采取了将影响国家教育政策制定的代表性人物,以及他(她)们在表述教育政策时所提供的代表性文本联系起来(比如:政协委员与提案文本,社会公众与群众来信,政府人员与教育政策等)。这样可以从文本的典型性和权威性表现有关教育政策的话语特征。这样做无论从文本收集的可行性,还是选取文本的稳定性方面都得到了保证,并且对研究方法的信度也可以重复验证。

按照上述与教育政策相关的主体人群合法地正式表述有关教育政策的言论途径,以及所呈现的教育政策的有关文本形态。围绕着政策主题,对文本的选定遵循了文本的制造来源是从组织理性出发到自发的(非组织理性)诉求的话语谱系分布排列,本研究确定了教育政策话语分析以文本为指向,根据文本与教育政策的关联性,主要选定了以下所列的五种文本作为能够表现教育政策及其相关话语的样本进行研究。同时确定了以"主题—文本—社会语境"为向度的分析框架,对教育政策进行主题话语分析。

本研究选定的文本有:

- 教育政策文献:中央领导有关教育的讲话报告和指示,教育部公布的教育政策文件(包括《教育部政报》),以及有关领导阐述教育政策的工作报告。
- 年鉴:记载教育发展和教育政策信息的年鉴,主要以《中国年鉴》、《中国教育年鉴》为主。
- 群众来信:群众就教育政策问题向教育部投寄的信件。
- 专家评论:专家(媒体)有关教育政策的评论和访谈录。
- 政协提案:政协委员有关教育政策问题的提案。

研究方法与策略

研究方法主要采用话语分析方法，根据教育政策所具有的特殊文化语用功能，本研究所选用的话语分析方法主要有以下三种不同的运用形态：

文本的主题分析

教育政策文本主题话语的确认与分析，文本采用了哪些主题词或主题话语概括表达了文本的哪些主要内容、揭示了哪些主要的政策信息、反映了政策话语变化的哪些语用特征。对于主题词出现的词频进行统计与分析。主题话语的问题域结构以及主题所涉及的知识谱系特征。此外还有有关政策主题的论说模式分析，表现主题的文体特征、叙事格式、说话人的社会身份分布特征、话语发生与地域经济特征之间的关联度分析。

文本分析

涉及教育政策的文本分析包括以下内容：

文本结构化分析和文本传播过程的叙事分析，按照语言学中语义、语用、语法的变化规律，将非结构化的文本转化为结构化的文本，进行语用特征分析（包括政策文本结构的分类表述以及关联的政策组合特征，即教育政策建议的文体风格）、教育政策传播特征分析。

文本的“互文性（intertextuality）”分析：即分析文本与教育政策主题之间关系的社会语言学分析，指教育政策以及与之相关的文本，群众来信、提案和有关教育政策的评论等文本之间的语义关系分析。话语技术化、话语秩序、话语知识谱系特征的分析（是聚合，还是离散）。

教育政策的社会语境化分析

语境化分析（contextualization analysis）包括对政策文本关联的社会环境、文化制度环境、经济环境等方面的影响要素分析，以及政策文本产生过程的社会背景分析，有关教育政策话语评论所表现的语言意识分析。政策行为中的决策意识分析。话语反映的政策变化机制分析等等。

在分析策略上采取了基于文本主题话语定量分析方法为主的研究,注意选取相同的时间间隔进行有关话语社会实践的时间序列比较分析。从中比较和发现教育政策过程中的内在运行机制。

教育政策过程中的内在机制分析

中国国家教育政策的发展过程中的阶段性规律,可以表现为一系列政策过程的内在机制在起作用,笔者认为,中国国家教育政策在重大战略型的教育政策决策和实施上,遵循着"全国教育工作会议"这种国家意识主导的"政府仪式"来倡导教育政策的推行,并且"全国教育工作会议"的召开是有规律的,这个会议的召开表现出国家重大教育政策从总体上遵循着"渐进层递的变化"模式进行,反映出中国国家教育政策决策的有限理性和精英决策特征。从世界公共教育政策决策的横向发展比较看,中国教育政策决策主体尚处在精英决策与共同体决策相结合的政策决策阶段。距离"共同体—公众参与"型决策和"公众尊法广泛参与政策决策"的制度境界还需要相当长的实践路程。具体从教育政策主体的波动周期变化,以及群众对教育政策执行情况的自发诉求话语,政协委员有关教育政策的提案建议都反映出教育政策决策过程的有限理性探索特征。这就是走"中国特色社会主义道路"的国家教育政策发展的阶段特征,进一步讲,中国国家教育政策的决策模式变化将与教育体制改革相互关联,将从政策变化与体制改革相适应的现实,向建立一个面向长远的有社会公众广泛参与的具有国家政治权威的教育制度转型,以教育制度和教育法制规范教育政策的决策是中国教育未来走向科学发展的必然选择。

教育政策主题话语信息传播规律分析

具体的教育政策是通过国家领导人有关教育的讲话,以及国家教育部颁布的一系列政策文件来体现的。具体的政策主题就是这些政策文本的主题内容。本研究按照编辑的时间"由远及近"对《教育部政报》这一政府政务公开的文本进行主题数量变化分析,将文本的索引目录进行排列统合,形成一个"索引文本"。这样做是基于教育政策文件在写作格式上具有的两个文体特征:首先是文件记录的真实性和客观性。正如教育政策的定义所言,教育政策内容在教育管理系统中,属于一种管理操作的规则和要求,政策内容

是有关部门执行政策的行动指针。按照“政务公开”的“指导性”原则，公开发布的教育政策文本，基本上能够完整地反映所涉及的教育政策文件的原始语义。其次是文件标题与内容之间保持语义一致的关系。按照行政公文写作中“一事一文”的政策表述原则，《政报》中每一个文件都要求表达清楚一个主要的政策主题，必须采取“事由—事项—结论”三段论式的“主题—述题”写作格式，准确传达相关的教育政策信息。因此，可以认为政策文件的题录数量与具体政策信息内容的增长之间呈现一致增长的变化关系。

基于以上前提，以及综合考虑研究文本的适用性、研究时间安排上的可行性，以及作者个人的知识能力等多方面因素，选定从1989年以来的《教育部政报》（简称《政报》）作为主要的分析文本，同时将相关时期的《中国教育年鉴》（简称《教育年鉴》）、《中国年鉴》等有关的政策文献作为分析所参考验证的文本①。进行这样的文献统计计量，是为了对来自政府已经公开的教育政策文本总体发展有一个比较精确的了解，同时对于《政报》的目录分类分布的状况有一个比较全面的掌握。

在统计和分析方法上采取了以下步骤，首先将选定的《政报》文献按年度规整，做出每一份政策文件的题录和主题综述，这属于文献计量的前期阅读准备工作。接着对《政报》按照年度和题录的分类项目进行记数统计，数据汇总。进行《政报》文献目录自然分类结构变化研究，为随后将原有的20类教育政策文本分类统合成7类提供数据规整的依据。接着采用文献传播计量学中有关文献老化率的计算方法建立中国国家教育政策的话语信息传播模型，并参照《教育年鉴》进行政策传播的影响因素分析。随后，采用信息传播学中的波动性测度计算方法分析各类教育政策在主题话语传播方面的相互关联关系，参照《教育年鉴》内容来分析影响传播的外在因素。最后对于分析的结果进行归纳和总结。从文献信息传播学的角度看，教育政策的过程就是有关某一主题话语信息在传播中不断衰减的过程，也是新的政策理念在不断替代旧的政策理念的过程。

政策的制定者在制定一项政策时，不单要考虑政策的价值和影响力，还必须对其颁布的政策存在时限有一个清醒的判断，政策存在的时限就是一个有关政策信息的老化和消失的问题。政策信息

---

① 有关国家教育政策的正式记录文本形式很多。主要收录各个时期政府有关领导人有关教育政策的讲话、报告以及政府有关教育政策的文件。

老化的时间特征对于判断政策传播的关键时间具有重要的作用。为此,作者对《政报》从 1989 年到 2002 年期间所收录的文件,按照《政报》目录索引进行文件数量统计,为了与 20 世纪 80 年代教育政策的变化相比较,我们将收集到的 1979 年、1984 年、1985 年和 1986 年的《教育部教育法令法规文献汇编》所收录的文件数量也进行了统计,作为分析的参照数据。以上两个时间段的文件按照年度汇总排列统计的结果如表 9.1 所示:

表 9.1 《教育部政报》按年度汇编的教育政策文献数量变化统计

年份	1979	1984	1985	1986	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
篇	123	180	187	239	133	284	310	263	261	165	175	176	133	155	167	172	171	179

注:①"年份"表示《政报》发行年度,"篇"表示每年度《政报》收录的教育政策文献数量。

②表中数据来源:a. 国家教委编辑《国家教育委员会政报》(1989—1999 年各卷);b. 教育部编辑:《教育部政报》(2000—2002 年各卷);c. 教育部编辑:《教育部教育文献法规文件汇编》(1979 年卷,1984 年卷,1985 年卷,1986 年卷)。

将上述数据,按照文献科学计量学中有关文献增长测量的计算方法,以 2002 年作为统计的起始年份,采用《政报》发行时间逆推回溯累积并求解出相应年度文献积累的百分比。这种"回溯式"的推算是文献计量研究方法中,为测算文献信息老化规律必须特定使用的统计算法。这里,作者根据文献计量学中有关文献老化的统计方法,以及"伯顿—开普勒法则"(庞景安,1999:139)的基本条件要求,首先验证了《政报》收录的文献分布符合文献计量学中的布拉福德统计分布特征。因为政策文件的归档具有严格的程序,《政报》的出版具有固定的时间要求。因此,可以认为教育政策文本作为一种政治学科的文献,也存在着一个政策文献信息老化的问题。在统计教育政策信息老化的趋势中,假定每份政策文献只按照被引用一次计量。由此,确定时间回溯文献数量积累增长率作为教育政策文献信息老化分析的测度指标。我们假定《政报》中每一份教育政策文件都表示一次政策话语行为(这是由政策文件拟定的"一事一文"撰写原则所限定的),每一个年度的政策文本的制造和传播都考虑了上一个年度以及历史上相关题材的政策内容,这样,教育政策传播的过程就是一个政策话语信息不断积累的过程,教育政策话语就是在信息不断积累的过程中不断衰减的过程。教育政策的传播过程可以表示为:其中第 i 年度教育政策信息的积累率用 Y_i 表

示,这样国家教育政策的信息传播可以按照年度的回溯,描述为以下的信息积累率数列过程:

$$Y_{2002}, Y_{2001}, Y_{2000}, \cdots, Y_i, \cdots, Y_{1984}, Y_{1979} \cdots$$

根据文献计量学中有关文献老化测度理论中"伯顿—开普勒法则"的基本条件要求,笔者提出采用政策信息传播中的信息数量积累增长率作为描述教育政策传播特征的政策实验测度指标,建立国家教育政策话语信息的传播实验模型如下,教育政策的话语系统特征为 P,系统特征的模型可以概括为:

$$P = F(Y, A, H, i, n)$$

其中,F 表示 Y, A, H, i 之间的函数关系。这个模型表示教育政策话语系统的特征(P)与政策信息积累增长率(Y)、政策文献的数量(A)、政策传播的方式(H)和时间(i)以及相邻年度之间隔数(n)等的变量相关。

政策的话语信息总量为:

$$Q = \sum_{2002}^{1979} (Y_i + S_i)$$

其中:S_i 表示有关教育政策话语信息传播衰减程度的积累率。

从信息传播衰减的角度,具体的教育政策传播特征可以更详细地表达为以下的模型:

$$S_i = \frac{\sum_{2002}^{2002-i} (A_{i+1} \cdot H)}{\sum_{2002}^{1979} (H \cdot A_i)} \times 100\%,$$

其中:i 取值范围在(2002,2001,2000,…,1989,…1979),A_i 代表 i 年度《政报》收录的政策文献总数。

H 代表教育政策话语传播方式的动态特征函数:

当 $H = e^r$,且 $r = n, n-1, \cdots, 2, 1$ 时,表示教育政策话语按照几何级数链式传播(几何级数传播模式)。

当 $H = r = n, n-1, \cdots, 2, 1$ 时,表示教育政策话语按照线性匀速递减形式传播(线性传播模式)。

当 $H = r = n = 1$ 时,表示教育政策话语按照原始文献数量保持原始形态传播模式。

本章中 n 表示从 2002 年回溯的各相邻年度之间的间隔段的序

数值,n 最大值取 17。

用表 9.1 中《政报》的文献数量值,按照不同的传播方式的动态特征函数,代入模型式分别求出时间回溯递推的文献数量累积之和,然后再求出每个年度对应的政策文献信息传播情况的模拟信息积累增长百分率,以 2002 年作为时间回溯的起点,当年的信息积累增长率设为零。按照年度时间回溯情况对应做出变化趋势图,就得出一个从 2002 年到 1979 年有关政策文献信息传播积累率随时间回溯增长变化的趋势图(如图 9.2 所示)。

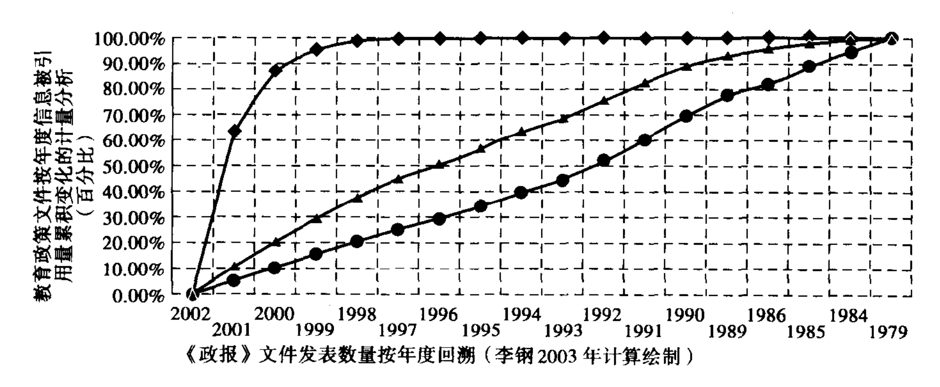


图 9.2 《政报》反映的教育政策信息传播减弱情况随时间变化特征分析图示

从图中可以看出教育政策文件的数量积累增长大约在 10 年内递增 50%。在文献计量学中,伯顿和开普勒(1960 年)把科学文献的半衰期(half-life)定义为"发表所有最新被引文献回溯累积量达到一半的时间"(庞景安,1999:145)。并提出了"伯顿—开普勒法则"概括了有关文献数量增长与文献内容信息的衰减之间的变化规律关系。"伯顿—开普勒法则"的推论认为"文献数量的回溯积累增长与文献信息衰减速度成正相关变化趋势"①。本研究所做的这个变化趋势图的曲线走势符合上述法则的推论描述。由此可以认为,《政报》中教育政策文件的数量积累过程中,有关政策内容的

---

① 上述公式中 A_i 是 i 年度《政报》收录的政策文件总数。i = 年数。另外,按照文献信息积累的计量统计规律,当 i 值足够大时(比如当 $i \geq 50$ 时,即半个世纪,是一个重要的文献统计时间计量时段),新旧文献在被引用的文献频次方面所表现的数量积累差距将拉大,整个统计变化趋势曲线将呈现负指数分布,50% 点将向起点收缩,文献信息的半衰期缩短。也即所谓知识更新的速度是与知识数量的积累增加之间呈现正相关变化的趋势。在文献计量学中有关文献老化率统计规律的积累测度模型是:$F(t) = 1 - e^{-bt}$,这里 t 表示从现在往回溯的时间年份,$F(t) \in (0,1)$,它表示文献积累增长的百分率。有关文献积累率的科学统计推算方法,请参见:庞景安,1999:123-170。

信息随时间变化发生了信息的新陈代谢，即新的教育政策内容在不断取代旧的教育政策内容。

综上所述，国家具体的教育政策话语信息传播整体上以"年"为单位。对于教育政策话语信息传播的分析，是为了认识政策话语信息变化的总体规律。事实上的国家教育政策内容更为复杂，政策话语不可能会理想化地仅沿着上述一种途径来单调传播，而是表现为三种传播模式相互交织组合形成复合的传播势态。这其中政策传播渠道的多种多样、传播的速度快慢不均，都决定了教育政策话语传播从总体上是属于信息非线性传播形态。尽管如此，上述实证统计所做出的"弓弦模型"中的那个"弓"与"弦"所围成的"半月型"信息衰减区域还是带给我们一个比较形象的概念，那就是政策信息的衰减具有一个确定的时段。这就是说，从一项教育政策颁布后的第 1 年到第 10 年，是整个教育政策执行的关键时期。具体哪一项政策在什么时间开始信息老化，需要采取怎样的措施来加强，还要根据具体政策的情况来定。从本研究的测算结果看，其信息老化的大致时间不会超过 10 年。任何教育政策从制定公布的那天起，就进入一个被社会广泛引用的传播速度增长时期。这期间社会有关教育政策的多种话语在共鸣，政策信息在迅速传播，直到教育政策为社会普遍知晓以后，就已经进入传播速度下降的信息老化时期。这就是政策发展中观念影响和政策话语消长的辩证发展规律，没有永恒不变的教育政策，只有永久传扬的教育话语。

教育政策执行过程中的群众自发诉求意识分析

在现阶段，中国国家教育政策的执行过程往往伴随着社会公众、学校和政府部门有关教育的利益调节过程。社会公众有关教育政策执行绩效的自发诉求，将通过寻求合法的行政救助的程序来解决。公众来信和上访等均属于国家政治意识形态确认的行政救助方式。公众来信中的自发诉求机制反映了个体与国家教育之间的政策调适关系。首先，公众来信总体上体现了一种"持续关注"基调话语的特点，以维系公众话语谱系的连续性。同时，又从较广泛的基调话语中蒸发出少量热点主题话语，表明公众来信话语谱系是收敛型的，即公众来信话语谱系发端于现实问题意识，在民间实践层面发散，随信访时间序列，终端被政府所收听到，并且部分可能进入政策问题范围，逐步收敛于决策层面。虽然，教育研究的学术语言谱系本质上是发散型的，但仅就教育政策问题而言，也可能是迅

速收敛型的,因为专家学者凭借专业话语权势的影响力,有可能实现从研究者到决策者的直接对话,促成实际教育问题直接反映到决策层面。公众来信话语谱系与学术话语谱系的收敛特征的差异,不仅体现在公众来信与专家学者在知识领域和信息依据方面的差异,更主要的是,公众来信关注的是现实教育中与切身利益相关的感性知识领域,而专家学者更为关注的是能触及教育系统规律的理性知识领域。

社会公众来信通过对于教育政策的自发诉求机制,所体现的是一种写信人(个体)与国家教育政策之间就受教育的权利进行的话语调适关系。对众多来信的阅读可知,众多来信的话语似乎都在"不谋而合"地述说着一个共同的主题,这就是来自社会最基层的人们对于国家教育政策"真诚的"关注、信任与利益期望。引发这种关注行动的是社会公众当中存在的一种自发的诉求机制。来信的话语秩序性以及向"技术化"方向建构的趋向,表明这种机制不仅具有"自发性"而且也具有某种"自学习"的社会化功能。话语谱系则表现了这种机制在行政制度体制中所形成的一种社会话语活动的"踪迹"。这种自发诉求机制所体现的是教育政策指向与社会公众接受教育权益之间的调适关系。这种调适关系反映了教育政策在制定过程中并没有完全搬用来信中所提出的公众自发诉求,而是慎重选择了来信中提供的社会事件"素材"作为国家教育政策决策所参考的信息资料。

正如上述所言及的,来信话语在维护国家教育管理的正常运行中,发挥着一种"安全阀"或"晴雨表"式的"问题信息提示"作用,教育政策的制定仅仅是将公众来信话语作为了解认识民意民情的一项政策调研途径。实际上在制定教育政策时,形成政策议题的话语构成成分复杂,往往汇集了影响教育发展的各种政治力量的利益诉求,最终所形成的政策文本是经历了一系列的政治协商谈判言语行动所达到的话语均衡文本。对于本章分析需要说明的是,限于公众来信样本数量偏少和可用的分析方法过于分散等因素,我们仅能以公众来信分析为主线,选择了话语秩序及技术化程度、语篇中的问题域结构和话语谱系等方面,来初步分辨公众来信的话语特征,并在此基础上分析与专家学者话语特征方面的差别。对照同期的《教育部政报》可以看出,由于公众来信话语谱系中浓重的感性色彩和随意性的语体,与政府严谨的政策文本距离甚远;同时,专家学者话语谱系中纵横交错的逻辑推证,与简明的政策文本风格也不合

拍。因此,在常规条件下,政府不可能仅依据某方面话语谱系匆忙做出针对性决策。在教育决策科学化和民主化的进程中,政府在关注公众来信的问题诉求和专家学者的问题争论时,政策文本在这里更像是一个"均衡器",将在教育政策问题认定和决策酝酿中,谨慎地选用和参考两方面的话语。具体的教育政策行为在国家教育行政部门与社会公众要求平等接受教育机会的关系中,扮演的是一种利益的均衡调适作用。

我国实行人民民主专政的国体,这从制度上决定了公众依法自发有序地采用信访等言论形式参与政治的合法性。群众来信也历来为我党和政府所重视。中共十六大报告提出:"改革和完善决策机制。要完善深入了解民情、充分反映民意、广泛集中民智、切实珍惜民力的决策机制,推进决策科学化民主化"以及"建立社情民意反映制度"。有学者从社会行为学理上认为"信访是公民参政议政的特殊通道,也是弱势群体维权的有效法宝;信访是民怨的释放通道、民情的反馈渠道、民声的传声筒、民意'用脚投票'的表达机制。"作者通过对信访行为的总体研究认为,信访和上访从根本上表现出民众对于国家教育政策和法制的无比信任和尊从。对公众来信话语特征的分析,有助于促进管理层对公众信访中暗含的深层次话语意识的重视,进一步提高决策层对相关教育热点问题的洞察水平,选择更加合适的决策模式来应对挑战。在决策的过程中,政府对来自一般写信者和资深学术专家的观点权衡方式可能有所不同,但是,有三个共同特征需要政策决策者注意,一是应当对公众来信中发生频率高的、公众反映强烈的教育话题进行认真的甄别和研究,对这些话题能否进入政策议题也可组织专门的论证。二是应当进一步建立和健全政府信息公开制度,保障公民的知情权,促使公众掌握更为充分的信息,以提升来信话语参与政策决策借鉴的技术化水平。这是政府"建立科学民主的决策机制"、"依法行政"和"民主监督"的重要环节。三是加强对信访信息接受与分析解决的综合联系,建立群众来信信息回馈机制,以体现政府勤政为民的办事效率,树立政府的公正信誉形象。党的十六大以后中央确立了"执政为民"的政府管理原则,中共十六届三中全会提出"树立全面、协调、可持续的发展观"的要求,以及 2003 年 12 月的中央经济工作会议更令人瞩目地提出了"以人为本"这样极富人文主义精神和现代执政理念的新思维与新方略。中国政府的国家政策发展决策观念正在向人本主义复归,这是顺应人类主流文明发展规律的"倾听·

牵挂·嘱托”和“民本”之举。具体到教育政策的决策机制建设上,一定要将信访信息分析与政府决策的信息采纳机制真正切实地联系起来,摸准中国教育发展现实的真正脉动,建立适合中国教育发展国情的教育政策问题信访反映机制,将教育政策在执行中的问题及时有效地控制在政府公共管理的有效解决框架之内,这是保证公共教育政策能够更有效地调整和促进我国教育改革与发展的一项非常重要的制度建设工作。它将有力地推动中国教育政策走向新的发展境界。

教育政策决策过程中的触发机制分析

人民政协委员的提案对于教育政策表现出一种“持续的政策参与”和“话语咨询”。反映出国家政治协商、民主参与制度框架中,开放式的建言机制。这种机制的开放性表现在提案的程序和话语建议规则设计上,对于提案者的发言机会和时间是多样化的和均等的。每一位委员同时可以有 4 种途径将提案反映到教育政策的决策层面。同时,提案的办案制度是开放式的交往互动模式,“案案有答复”的落实制度。这是依法治国所必须的制度保证。

提案的主题话语和问题域在前后两年的相关稳定变化势态,体现出一种制度稳定的有限理性主义话语倾向,具体讲就是对于教育政策的评议话语变化是以渐进式的发展演化为主导的。新的政策话语的产生是建立在渐进的话语基础之上的发展样式。这种渐进的教育政策语境与教育政策变化的“渐进层递模式”不谋而合。决定两者内在关联的话语意识在于,国家教育政策的决策一直处于稳定的政治发展语境当中,这种语境内涵的价值观提倡决策的制度有限理性主义。所谓制度有限理性主义是指,人在制定教育政策时是理性化的,或者说是想要理性化决策的,但同时所拥有的理性是有限的。因此,就有关教育政策的整体设计和制度设计,我们难以完成全面完美的设计;但对于局部的教育发展战略,却完全可以通过科学的论证和努力“试错”、汲取前人的经验,来开发和学到设计合理的政策的技术和方法。这种制度有限理性主义在提案行为中反映出对教育政策的不断总结、评议、观念话语建构与新政策的设计。例如,2001 年和 2002 年的提案中都有要求对 1999 年开始的“高校扩招”和“院校合并”政策进行有关方面的评估,这些评估集中在教育规模发展过快以后,人才培养质量与表现效益之间关系的衡量与评价,并就此提出了一些兼顾教育发展速度与质量平衡协调的政策

建议与政策设计。

民主政治的改革与进步，不断丰富民主参与政治意识的语境。按照列宁的国家学说，民主首先反映为一种国家制度。“民主是一种国家形式，一种国家形态。”“民主意味着在形式上承认公民一律平等，承认大家都有决定国家制度和管理国家的平等权力。”（列宁，1985：257）这种建立在国家制度上的政治民主平等意识，在教育政策制定中的反映就是要提倡“教育公平”、“人人都有受教育的权利”、“技术和职业教育应普遍设立，高等教育应根据成绩而对一切人平等开放”等公平的教育政策决策价值观念（联合国教科文组织，2000：16）。邓小平就有关民主参与政治很现实地指出，“社会主义民主政治是实现四个现代化的政治基础”，“没有民主就没有社会主义，没有社会主义就没有现代化。”强调要加强民主参与意识，“发展和完善社会主义的经济制度和政治制度。”党的十六大报告将“发展社会主义民主政治，建设社会主义政治文明”确立为社会主义现代化建设的重要目标。政治环境和民主参与政治制度的发展都在提案内容的变化上得到体现，形成了良性合法化的民主参与教育政策决策语境。阿尔蒙德认为“参与型的政治文化更易于动员人民”，“只有参与政治的合法化，才能够为公民提供参与决策过程的机会，这有利于了解、信任、支持政策，约束非理性情绪化的政策判断。”（阿尔蒙德，1987：300）提案话语主题分布的广泛性以及提案人社会身份分布的集团化的特征在暗示着，这些来自社会不同阶层的利益代表对教育政策众说纷纭的话语中，存在能够影响教育政策决策的“关键话语”。是哪些群体的话语真正触动了教育政策制定议程的“开关”，是怎样的话语将日常生活中的教育问题提升到教育政策的议程当中？这其中就涉及有关教育政策制定的“触发机制”概念。

拉雷・N・戈斯顿（戈斯顿，2002：117）有关公共政策发生机制的研究表明，在开放的社会环境影响下，客观存在着一个有“规律的”“触发机制”，促动公共政策系统发生变化，诱发政策议程启动，制定出新的政策。触发机制是诱发公共政策产生的“导火线”。触发机制由“触发事件”和一系列导致政策系统发生变化的话语传播方式所构成。具体讲就是，先有一个（或一系列）重要的事件发生。整个事件把例行的日常问题转化为一种引起普遍共识的、消极的公众反映。公众反映对政策形成压力，成为政策问题的认识基础，通过诱发因素作用，公众话语随之引起触发事件，将政策问题复杂化，

政策决策者要应对复杂化问题,就得将政策问题转化为要求政策制定的动力,政策议程随之启动,并且会根据触发机制的时间持久性而发生质的改观(新政策浮出水面,以应对触发事件)。托马斯·戴伊(Dye,1987:55-59)对此评论道,政策活动"受到环境性质的极大影响",触发机制有助于定义政策问题以及随后的政策选择。当提案话语主题符合触发机制的条件时,就可能借助触发机制,成为教育政策的内容。这个触发条件就是:触发范围、触发强度和触发时间。作者将上述有关触发机制的理论拟采用程序化的"决策树"语句叙述如下:

触发机制的发生就是要:

- 有符合触发条件的触发事件发生。
- 事件结果导致产生压力并集中作用于政策系统的某一环节。
- 政策决策系统为了排除压力、维系正常运转,要将面临的压力问题列入政策议程。
- 需要制定新政策以应对压力或消除压力。
- 新政策产生。政策实施产生效果的途径有两条,一是政策有效,政策调控的系统向政策预设的目标发展。二是政策失灵,系统向不利的方面转化,形成政策危机,诱发出新的更大的触发事件。

其中触发条件也就是导致触发机制生效的关键要素。这三个要素表现在政协提案上就是:触发范围,指提案人数和有关主题的提案发生频数的众与寡;触发强度,指提案者或集团的政治影响力大小,以及有关提案在政策决策者层面的影响力程度;触发时间,指提案和有关的政策建议提出的时间是否处在一个集中政策决策者关注、社会教育问题的压力发生、学校发展的需要三方面博弈的信息交汇"关键时段"。

由于篇幅所限,作者用2002年全国职业教育工作会议的召开,国务院印发《关于大力推进职业教育改革与发展的决定》作为具体解释触发机制的政策分析案例。

〔**背景叙述**〕职业教育是改革开放以来国家一直重视的领域。1995年5月15日《中华人民共和国教育法》颁布,翌年,《中华人民共和国职业教育法》颁布,这个法对职业教育的办学原则、教育方针、职业教育发展与科技生产力和经济发展的关系做出了法律层面的规定。随着我国经济建设和社会发展,

尤其是社会对职业技术人才,特别是第三产业从业人员的需求急剧增加。进入新世纪以来,中等职业学校在校生占高中阶段教育的比例出现了一些变化,1993 年这个比例数值为 54%,到 2002 年下降到了 42%。社会要求进行职业教育管理体制改革的呼声强烈。从 2000 年开始,每年"两会"上,都有人大代表和政协委员就有关职业教育进行提案和政策建议。

〔**触发事件发生**〕尤其在政协提案方面,2001 年到 2002 年职业教育的提案发生频数逐年上升。到 2002 年有关"职业教育"的提案内容成为名人提案、党派提案、团体提案等"个人—集体"提案共同关注的焦点。一些具有政治影响力的民主人士积极呼吁国家要召开"职业教育工作会议"专门研究职业教育的发展问题。诸多提案要求职业教育改革集中在两个方面:一是教育质量,二是学校内部办学管理体制改革。党和国家领导人出席了 2002 年的政协会议,参加和听取了有关委员就职业教育的大会发言和小组讨论。2002 年 7 月 8 日,国家科教领导小组第 11 次会议召开,研究加快职业教育发展和实施现代远程教育工程问题〔**以上构成了有关某项教育政策的触发事件,以及触发范围、触发强度等要素**〕。

2002 年 7 月 28 ~ 30 日,国务院在北京召开了全国职业教育工作会议,8 月 24 日国务院印发了《关于大力推进职业教育改革和发展的决定》,确定了职业教育要重点推进职业教育管理体制改革、办学体制改革和教育教学改革,职业教育学校人事制度与就业制度改革,促进就业与再就业的迫切需要。

综上所述,从政治协商、参政议政、民主监督的语境看,触发教育政策制定的三个要素在提案与教育政策制定的关系中表现为话语范围,即参与某项提案的人员的普遍性。话语强度在提案中与进行教育政策建议的群体的政治影响力,以及提案人所代表的社会团体的经济影响力有关。来自政治和经济方面的影响力将给政策制定者带来政策决策的压力和决心。在有关触发时间方面还应当考虑到那些触发政策的话语事件发生的社会环境和传播速度的快慢是否"适合时宜"。例如,汤敏博士提出扩大高校招生规模的政策建议,在当时的经济发展环境、就业形势和教育发展阶段就"适逢其时",所以,他的政策建议很快就被采纳作为 1999 年全国教育工作会议的一项政策,迅速地传遍全国。从时间上看,传播速度显示着教育政策的力量。

提案的话语还反映出民主参政制度设计与教育政策发展之间的关系。实质上是指社会群体利益与教育政策发展之间的关系。比较政治学的研究认为:可行的民主重在广泛参与的话语实践行动表现。“人民当家作主”就表现在社会公民能够通过特定的机制,最大限度地实现自主治理,并有效地使官员对公共利益负责(毛寿龙,2003:43)。根据这一知识,政治制度才能够被设计成现实可行的民主参与形式,并有可能防止民主被别有用心地“滥用”。此处所言的“民主参政议政”以及“自主治理”是与政策的设计相伴相生的。参政议政民主化是测度政治发展的重要标尺(杨帆,2003:35),也是检验教育政策制定水平的重要尺度。阿尔蒙德(1987:89)认为:政治发展首先是政治文化的世俗化,即某种政治文化成为普遍化、大众化的文化,还有政治角色结构化和自主化,最终是为提高政治决策能力和贯彻政策的能力。政治体系扩大的同时,其政策能力要相应增强。提案与教育政策制定之间的关系也是如此。只有选择大众化的教育政策,才能满足全社会对于教育的普遍需求,才能促进教育发展真正符合社会大众文化知识增长的需要。只有对绝大多数国民带来利益产生影响的教育政策,才有可能成为促进国家和社会发展的教育政策。

就教育政策议题的话语产生来看,从提案方式构成中“集体形式提案”的比例占到总提案数量一半以上,就足以证明社会对教育政策以“集团化”建议将继续主导着教育政策发展的议题话语来源。教育政策的制定目前仍然难以超越来自城市中的某些特定知识阶层的话语影响。这反映出一种具有现今中国国情特色的影响教育政策决策的重要社会文化意识语境制约。从提案反映的政策问题目前主要所关注的范围集中在“教育质量”和“教育投入与效益”这些方面,同时伴随着有关“加强德育”、“素质教育”,提倡“受教育机会平等”的利益呼声。国家教育政策在新世纪以来的一系列重大举措,反映出公民政策参与和教育政策发展的同步进展样式,尽管这种进展样式在目前仍然难以逾越现实经济条件和社会发展的客观环境制约,但教育的崇高理想会始终贯穿于教育政策的行为当中。在面向大众化的公共领域中,促成教育政策的话语在不断地努力摆脱经济利益语境的纠缠,包含来自社会更多阶层的文化内容,调整城市与农村之间的教育差距,重新回归到传承人文精神和知识文化理想的教育“本真”语境当中。指导教育政策制定的价值观念中非常需要加强人本主义精神理想与信念的培育。

表 9.2　社会各界委员提案呼声的案例数量比重变化表

提案的主题话语比重	社会各界委员以个人形式就有关主题的提案情况（个人提案占总提案频数和的比重）										
	高校界	军队界	科技界	民办学校	民主党派	企业界	社会界	文化艺术	政府界	中学教师	宗教界
2001 年的比重数(%)	30.30	1.14	5.30	1.14	7.95	7.58	13.64	9.47	17.05	3.03	3.41
2002 年的比重数(%)	32.28	0.00	3.17	2.12	12.17	10.05	11.11	8.47	20.63	0.00	0.00

注：本表仅采用以个人提案的形式做统计，其中的提案呼声案例数量比重 = R/Q；此处 R 代表某一界别委员个人提案频数合计，Q 代表相关年度所有界别的委员就教育政策个人提案的总数量。李钢 2004 年 1 月制表。

表 9.3　2001—2002 两年度对教育政策提案的政协委员社会界别分布一览表

社会界别	2001 年提案频数	占年度比例	2002 年提案频数	占年度比例	两年合计	所占总量比例	总排序
高校	76	40.86	43	42.16	119	41.32	1
政府	27	14.52	25	24.51	52	18.06	2
民主党派	16	8.60	12	11.76	28	9.72	3
文化艺术	13	6.99	1	0.98	14	4.86	5
科技界	9	4.84	3	2.94	12	4.17	6
医疗卫生	3	1.61	2	1.96	5	1.74	8
社会	17	9.14	6	5.88	23	7.99	4
中学	5	2.69	0	0.00	5	1.74	8
军队	4	2.15	1	0.98	5	1.74	8
小学	1	0.54	1	0.98	2	0.69	10
企业	9	4.84	1	0.98	10	3.47	7
民营企业	1	0.54	4	3.92	5	1.74	8
公益机构	1	0.54	0	0.00	1	0.35	11

续表

社会界别	2001 年提案频数	占年度比例	2002 年提案频数	占年度比例	两年合计	所占总量比例	总排序
专科学校	0	0.00	2	1.96	2	0.69	10
民办学校	2	1.08	1	0.98	3	1.04	9
宗教界	2	1.08	0	0.00	2	0.69	10
合计	186	100.00	102	100.00	288	100.00	

注:表中统计的身份是按照委员所来自的社会界别分大类进行统计,对于身份和界别不清楚的都归类在"社会界"这一类中。李钢 2003 年 12 月制表。

理解和认识新世纪中国国家重大教育政策变化的规律

社会话语镜像反映的中国国家教育政策发展状况,从总体的发展趋势上看,表现为在"激进—渐进"的教育政策均衡调整中追求"渐进和谐"的教育发展战略。这种渐进和谐的国家教育发展战略可以从以下两方面进行话语考察、理解和认识。

影响国家教育政策发展的文化变量

对教育政策的不同话语评论,表现出人们对于教育政策所持有的不同价值观。这种价值观又形成一种文化氛围,决定着人们参与教育政策评论的话语形式,影响着教育政策作为一种政治文化进行传播和信息交流。在就有关教育政策的话语系统的文本内容选择机制上,存在着不同社会身份的人群选择政策评论话语形式时具有不同层次的言语合法性。教育政策作为国家和政府组织有关教育事业的行动准则和文化规范,在教育政策相关的话语系统中,政策的决策者、参与者、研究者、传播者,以及教育政策的关注者位于社会的不同阶层,由于社会各阶层的人们与教育事业的关系和利益不同,在对同一教育政策的主题判断和价值认同上存在着客观的认识差异,他(她)们对教育政策的关注点、关注程度不同,表现为不同的话语形式。这体现了不同社会身份的群体有关教育政策的不同利益诉求和价值判断。从政策传播的过程看,本研究的写作所采用的是贯穿于"教育政策传播—社会公众反应"这条主线的政策话语

分析。

对于分析所用文本的特征认识,采用政策话语评论的理性化程度以及对政策认识动机作为划分话语类型的认识标准,结合说话人的话语行为方式是自发的还是有组织的,可以建立一组包括四个相互对应的短语形成的认识维度,在这种认识维度中,本研究选取的文本被相互联系起来,成为一个面向教育政策实际语用的话语体系。这四个维度概括地分为:

1."非理性话语—理性话语"维度:这里的理性是指组织理性,是组织利益和价值观的代表体现。理性话语是指话语的产生经过了一个程序化的"证伪"过程,证伪的语义是话语经过了组织价值观的筛选和判断,话语内容的选择带有辩证的全面性,说话者的价值观保持了一定程度上的"组织理性价值",与非理性话语的产生条件相对立。

2."自我动机—社会动机"维度:社会不同阶层的人们对教育政策的诉求从心理上讲是依照政策给他们带来的利益和价值观感受来建构不同的动机选择并确定话语行动方式的。不同的动机所产生诉求的话语指向是不同的。自我动机意指言语者创造文本的动机出自于以自我利益为中心(如信访),社会动机是以社会群体利益作为话语指向的(如提案)。

3.对政策问题"倾向于感性认识—倾向于理性认识"维度:指文本中对教育政策问题阐释与表述的话语方式主要是发自个人情感,还是对政策问题进行比较科学的分析和调查的基础上,进行的科学理性思考。教育政策的传播形式和途径,政策接受者的知识水平与经济境遇,以及政策传播所处的社会文化环境,影响着人们对于教育政策问题认识的理性化程度。

4."自发的话语意愿—组织的话语意愿"维度:指话语的产生方式上,政策议题的产生途径是出自于内心自愿的言语行动,或是按照组织既定的程序规定制造产生的话语。群众来信、专家评论(新闻媒体评论)、政协委员的提案、政策文件(政报)在代表教育政策的文化特征方面各具有的话语特性,决定了它们在上述四个维度中的位置,这些文本排列的相互关系如图9.3所示。

上述文本围绕着某一些教育政策主题,表达了不同社会阶层对于教育政策制定的影响关系。这其中贯穿着一种有关教育政策诉求的个人动机与社会组织动机的整合过程,构成了从教育组织的理性话语(政策文本)到民间自发的感性话语(群众书信)的政策语言

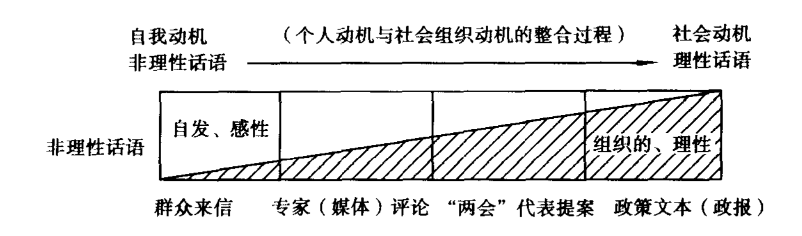


图 9.3 研究选取文本相互关系排列示意图

谱系。这个谱系是教育政策话语活动的主要文化空间。笔者认为，上述四个认识维度，在表述教育政策话语活动与政策变化之间的关联关系上，可以作为考查教育政策发展的文化变量。

确立一种"渐进和谐发展"的国家教育政策发展观

作者认为，国家教育政策发展观是指国家在适宜的社会环境下，适时制定出新的教育政策和新的教育发展战略，并使之发挥出有效的政策功能，来切实推进国家教育事业整体发展的政策决策观点。确立科学的教育政策发展观就是要让教育更好地为国家总体的发展服务，教育政策发展符合国家整体发展的需要。保持教育政策的连续性和持续发展作用影响，用政策指导教育发展，更好地满足人民群众日益增长的教育需求和受教育权益。

通过前面的分析可知，国家教育政策从总体上遵循着"渐进序贯"的修补机制变化。在国家教育政策具体过程中，来自不同社会阶层的话语功能在政策的不同阶段"突现"。在政策的制定和形成阶段，领导讲话文本确定了教育政策战略行动的核心话语，确定了教育政策的主要议题，体现了政策决策的集中制度和行动指向。领导讲话是政治组织集体精神的集中体现。

政协委员的提案是站在国家和合法组织所赋予的参政议政话语权利基础上，就教育政策的设计和决策中的问题进行有针对性地提问。他们关注的是政策制定与自身所代表的组织利益之间的关系、政策执行当中涉及群体利益的调整问题，其政策建议是带有社会阶层性的政治价值观倾向的。他们的政策评论话语是建立在构成国家意识形态的组织文化当中，对于教育政策的评论具有政治合法性，其政策建议是建立在教育政策决策信息更为完整的宏大语境中，其话语构成国家教育政策语义内容的主要来源，提案者也当然成为影响教育政策决策的主要社会群体。

在教育政策的执行阶段，群众来信按自发机制提出政策诉求，其知识背景处在不完全政策信息环境中的低语境状态，关注教育政策的具体执行问题与其自身利益之间的关系。群众来信的话语成为检验教育政策传播和执行情况变化的“晴雨表”。

通过前面各章的分析结论，从教育政策决策行为的内在意识上看，国家教育政策表现为对于个人、社会不同利益群体、社会发展与教育发展之间关系的调节。尤其在反映国家重大教育政策的变化方面，群众来信、政协提案和教育政策文件本身各自侧重于从不同的角度在对于教育政策作用与各自之间的关系进行解释。

改革开放以来的25年间，在中国政治发展稳定的大环境中，国家重大教育政策遵循着一种“渐进式的序贯层递模式”发展变化，并已经有效地调整社会不同地区的社会发展与教育发展之间的关系。教育的发展及时地服务于社会经济发展的需要，这已经成为教育政策文化功能的阶段性特征。

在教育政策的执行过程中，群众积极踊跃的来信体现出一种民间自发的教育政策诉求机制，来反映个人与教育政策发展之间的利益关系，要求教育政策能够代表基层群众的教育利益。在教育政策的整个行为过程中，政协提案的动议包含着有组织的促进教育政策制定的政策触发机制，提案反映了不同的社会集体对教育政策发展的要求。其中，受教育人口数量的增长，城镇居民的教育消费水平和参政议政者接受教育的水平是影响教育政策制定的外在条件。在教育行政组织内部，组织管理水平、政策研究基础则决定着教育政策的决策文化价值观取向。教育政策的变化还与教育行政组织机构精简、职能变化、组织整合以及功能专业化有关。总之，中国教育政策的变化仍然是来自经济、城乡人口结构和教育系统本身等方面的变量纷繁复杂地交织在一起互动作用的结果。

进行本研究的目标是想探究中国社会转型时期公共教育政策制定过程中的政策决策意识发展规律。正是由于制定政策的意识是难以观察到的，所以，作者从分析有关教育政策的文本和话语来“管窥”国家教育政策决策意识变化的“一斑”。由此所采用的叙述、语言意识水平测度、建立的话语秩序、采用的知识基础等都不外乎是从话语表达上在验证教育政策决策意识的存在及其变化的规律。这一决策意识变化规律首先表现在政府的教育政策文本上，自1979年党的十一届三中全会以来，无论是采用全国教育工作会议的形式制定和传播国家重大教育政策，还是采取《中国教育改革和

发展纲要》、国家教育发展"五年"计划、《教育振兴行动计划》等具体的战略型政策来规划教育事业发展,"教育改革和发展"这一话语始终是国家教育政策制定的决策意识紧密围绕着的主题。不断地进行战略型的教育政策设计,体现出一种渐进式的政策发展意识。

教育政策要指导和服务于教育改革发展,就需要坚持确立政策发展的意识。政策发展就是指要学会采用战略型长期政策来逐步有效地推进教育的持续发展。教育政策发展观代表了由个人、集体、社会政府对于教育改革在内容和行动选择上进行的一种目标共同的利益博弈。教育政策的内容就是这场博弈达到动态平衡时各方面的教育权利话语记录。尽管我国的教育在 20 世纪 90 年代中期就确立了"教育优先发展战略",1999 年开始又采取了"加快高等教育发展"的战略,但是教育是作为一个知识递进积累增长且文化内在关联性强的人才培养系统,动态平衡的调节规律一直贯穿于国家教育政策变化的过程。这就是为什么在过去的 25 年,国家教育政策从整体上一直保持着"渐进式的序贯层递"形态前进的内在原因。因为基础教育与高等教育两大领域之间在接受教育的人口规模、知识水平以及教育经费投入单位比重之间存在着的巨大差异,这种多方面的差异与教育整体发展之间相互牵制,从内部制约着中国教育跨越式发展的速度和步伐。在未来的岁月,尤其在 2020 年前中国社会发展的"战略机遇期",这种不同层次教育之间发展水平差异所产生的相互影响和牵制状况,将依然困扰着教育政策的制定者,减小教育差距是中国教育政策面对的重大挑战之一。在最近国家公布的《2004—2007 年教育振兴行动计划》中,将"农村教育发展"和"高水平大学与学科建设"同时列为两大战略重点(光明日报,2004-3-25)。国家开始采用"齐头并进"的均衡政策策略来尝试解决这一矛盾。

基于前面的研究,作者建议在国家教育政策制定中,要继续保持这种"稳中求快"、"渐进有序层递"的教育政策发展观。根据本研究前面所进行的有关教育政策的话语意识分析以及与社会经济的关系分析,作者认为在目前中国的政治经济和社会发展保持现在的这种稳定局势的前提条件下,教育政策应当着力于开放式的人才流动制度建设。如果不发生意外的社会突变,中国教育在新世纪将进入一个较长的稳定增长时期,这就需要推出新的战略来指导教育发展。作者认为在此期间促进新的国家重大教育政策出台,或者说

是促进国家重大教育政策发展的条件中应当包括以下五个方面的重要内容：

第一个条件是：要有一个符合触发机制的政策主题话语域。这是由教育政策所处的社会制度环境决定的。没有一个稳定和持续发展的政治环境、经济改革要素和社会各阶层对于教育的利益需求，教育以及教育政策发展就不能顺畅进行。从这个意义上讲，现今的教育政策发展确实是进入了一个"建国以来发展的好年头"。在科教兴国战略和人才战略的大环境下，切合稳定和利益和谐调整教育政策议题指向的社会舆情反映，就能够触发产生新的教育政策的新话语，新政策议题将不断涌现。

第二个条件是：调整个人、集体、社会及教育的关系发展与市场经济发展需求指向之间存在的差距。这是教育政策调整的条件，也是教育政策本身带有"反哺"功能的体现，教育依靠经济发展的基础来生长，又反过来作用于经济发展。教育政策在教育发展处于初级阶段时，主要着力与教育内部的关系调节，当教育发展进入与社会经济发展密切作用的阶段，教育事业已经成为一种生产力的培育基地，这时，教育政策的调整功能就要从教育系统的内部扩大到对于整个社会的影响。例如，"高校扩招"政策不仅仅调整了适龄青年进入大学学习的"毛入学率"，而且还间接地对于社会经济的财富增长与收益分配格局以及就业市场的波动变化周期进行了调节。在新世纪教育政策对于整个社会发展的作用将更加突显。

第三个条件是："生机勃勃的城市居民生活方式改观。"城市话语是目前诱发教育政策改革发展的重要条件。目前产生教育政策议题的话语发源地主要是在城市而不是在农村，是来自学校以及学校所在的社区。比如，教育政策的制定以及政协提案的大多数都是来自城市和高校。对于这方面的认识应当坚持马克思主义的历史唯物观点，实事求是地面对现实、分析现状。认识和理解教育政策，其目的是为了更好地给将来的政策提供更为积极的建议。作者认为农村向城镇化的发展步伐与人口向城市流动的社会变化，将促进教育政策向多元化调节的方向发展。促进教育政策将注重提高农村基础教育和职业教育水平，重视研究流动人口的教育政策，缩小农村与城市之间的知识差距。保持教育发展的平衡，提高促进国家经济发展的整体教育创新能力和综合竞争能力。

第四个条件是："党中央确立了一系列的重大政治方针以后……"这是具有中国特色的国家教育政策议程开始的序曲，是国家

教育政策制定的政治动力,是中国国家教育政策所代表的人民利益和政党政治性质的政治联系和表现。前面已经进行的有关中国共产党全国代表大会会议精神与全国教育工作会议精神之间关系的叙事分析,就是对于这种动力作用的一种解释。在20世纪的后期,国家已经召开了四次全教会,教育政策从宏观上已经经历了一个完整的政策变化周期。在新的世纪,党的十六大召开以后,一个崭新的教育政策发展机遇期即将开始,国家将有可能遵循已经运用娴熟的重大教育政策制定模式,在不远的将来再次召开全国教育工作会议,来制定和推进国家教育政策在2021年以前的发展战略。① 国家教育政策将继续弘扬"爱国主义教育"、"德育"、"素质教育"、"人才强国战略"等话语主题,坚持中华民族的团结和自尊自信、刻苦耐劳的优良传统,促进教育为国家的经济改革和政治发展提供更多的、合格的人力资源。

第五个条件是:教育政策的发展与国家对农村经济和教育问题的解决策略有关。无论是社会公众来信还是政协委员提案,这些来自社会不同阶层的代表与社会群众在反映农村教育问题的话语上,表现出"惊人的一致性"。中国的人口问题制约着教育发展的速度和水平,这早已是被证实的客观论断。中国的农村人口数量占到全国总人口比重的62.35%(中国教育与人力资源问题报告课题组)。人口问题的背后是经济贫困问题。实现"普九"、"两基"以及"农村税费改革中的教育管理体制改革"问题一直是近年来国家教育政策文件中反复强调的政策主题。农村教育问题是导致中国教育人口分布不均,受教育差距分布不齐,以及农村人口流动的主要原因之一。这其中根本的原因在于经济,或者说是"三农"问题。农村教育平均水平低是牵制着中国走向现代化的关键掣肘之处。农村教育是国家重大教育政策调整必须关注的领域。教育政策要发展就必须时刻关注农村人口受教育的问题。究竟是采用"文化普及型"的农村教育政策策略,还是输入先进文化,进行教育管理体制改革,采用"移风易俗型"的农村教育政策策略,作者认为两者应同时结合进行。但最终解决农村教育问题的关键出路还是在于农村经济走向富裕。国家教育政策不仅仅要推进农村全面实现"普九"和"两基"教育目标,更主要的是要通过教育政策进行农村教育综

---

① 这里所说的"2021年"仅代表本文作者个人根据本文研究所作的一种教育政策战略发展期的预测分析,所依据的来源是根据有关全教会的叙事分析时间段测度。

合改革，用教育来引导广大农民“移风易俗”，走上一条内含丰富物质文化内容和开放民主意识的现代化生活之路。同时要采取政策措施加强农村与城市之间的文化交往互动，通过政策指导，认真实施基础教育、职业教育和成人教育“三教统筹”，积极推进农业、科技、教育的“农科教结合”的经济发展战略（张力，2003：90-101）。通过“村村通”和“一村一名大学生”等教育项目。将先进文化传入中国广大的乡村，改造和阻止农村生活中的“愚昧习俗”通过人口无序流动“泛滥”和蔓延。采取综合的政策手段说服和积极鼓励支持乡村政权“资教、助教、办学堂”，以减少和消除农村教育与城市教育水平之间存在的知识差距，未来的国家重大教育政策仍将坚持教育平等的理念，致力于缩小城市与农村之间存在的越来越大的知识差距。

后 记

动手撰写后记时，已是2006年的正月初，清华园里飘扬着2006年的第一场雪，皑皑白雪漫漫纷飞给原本就青翠葱茏的水木清华景致更添了一份素雅宁静之美。“忽如一夜春风来，千树万树梨花开”，春天的脚步已经伴随着这场瑞雪悄然而至，我的心情也愈加爽畅舒坦。这本论文集就要在这场悠然漫舞的雪花和好心情陪伴中出炉了。

本书是我进入清华大学公共管理学院进行公共政策分析专职博士后研究过程中，自选的研究成果之一。编撰本书的目的旨在弥补博士论文研究期间，有关公共政策分析方法论研究的不足，同时认真反思目前中国公共政策研究领域普遍存在的“重观念宣扬，轻方法论专研”的缺陷，加强有关公共政策分析方法论问题的深度思考，以期取得有关公共政策方法论研究的阶段性研究成果。在本书的编写过程中，文献收集费用得到了中国国家自然科学基金面上项目资金(批准号为:70573059)的资助。于此，我特别感谢国家自然基金委对于中青年后进学术新人的热诚提携，愿国家自然基金委永远做培育中国科研创新新生力量的孵化器和温床。

有关本书所辑9篇论文的选取以及其中7篇文章的翻译版权许可谈判，得到了我的朋友正旅居美国芝加哥城的蓝石博士(Dr. Stan Lan)的大力帮助。书中所有英文论文资料的翻译出版授权许可均由蓝石先生义务帮助我与涉及原文的六位作者(或他们的委托人)逐一联系谈妥。值此文稿即将付梓之时，我谨向远在大洋彼岸的蓝石先生及他的全家致以诚挚的感谢。本书写作资料的收集整理过程，涉及到许多海外作者的协调，清华大学国际处副处长夏

广志老师给予我们在外事联络方面以非常专业的指导,在此特别表示感谢。

全书的编著过程比较妥贴地体现了中国的一首唐诗所叙述的高远意境:"海内存知己,天涯若比邻。"

若条件适宜的话,我将会诚意邀集全世界各国专业研究公共政策内容分析方法的专家相聚于北京,来个"华山论剑","他山之石,可以攻玉",跨国进行政策分析方法合作研究。不亦乐乎!

本书的外文翻译和校对工作由李钢博士承担完成。具体编译和写作的工作量为:第1章、第2章、第3章:李钢、邓春芳。第4章、第5章、第6章:江雪梅、李钢。第7章:李钢、徐志勇。第8章:李钢、江雪梅。其中第9章系本书主编李钢博士于2002—2004年间首次运用内容分析方法就当代中国教育政策发展独立所作的最新分析研究成果。于此我还需要特别感谢的是重庆大学出版社的雷少波和崔祝两位编辑对于本书的支持和学术价值的认可,他们与我第一次打交道在2005年9月间,在看完我所编撰的一叠厚厚的"草稿"后,就一致认定这本稿子里面肯定含有"玉"的方法论学术价值,并提出了非常详细的修改意见,一致同意出版。

由于全书是由一系列论文按照公共政策内容分析方法的运用为主线,从多种不同的文本编辑形成,其中难免还有一些整体系统上语义的不够连贯,那是主编的水平有限;其中一些作者基于不同的政治意识形态信仰,在分析观点上或许会有悖于目前中国学术界和公共政策学界习惯使用的逻辑论题,还请读者本着学习政策分析方法论为先的原则态度,在阅读本书时注意批判、辨析和体察。总之,让我们以积极建设的态度,为创新具有中国国情需要的公共政策分析方法理论体系进行有意义的和卓有成效的研究和开发,真正有效地服务于中国公共政策的科学决策,积极提供科学的政策分析方法理论支持。

李　钢

2006年1月修改于清华园

参考文献

[1] Airasian, P. W. (1997). Classroom Assessment, New York: New York, McGraw Hill.

[2] Babbie, E. (2004). The practice of Social Research (10th ed.) [M]. Wadsworth Publishing Company.

[3] Berliner, D. C., & Biddle, B. J. (1995). The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools. Reading, Massachusetts, Addison-Wesley Publishing Company, Inc.

[4] Burchell, G., Gordon, C., and. Miller, P. (1991) (Eds.). Politics and the Study of Discourse.

[5] Corson, D. Discourse and Power in Educational Organizations. 1995.

[6] Croker, L. M., & Algina, J. (1986). Introduction to Classical and Modern Test Theory, Orlando, Florida: Harcourt Brace College Publishers.

[7] Dye, T. R. (1987): Understanding Public Policy, 6th ed. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[8] Fairclough, Norman (1995). Critical Discourse Analysis: The Critical Study of Language. London/New York: Longman.

[9] Foucault (1977), M. What is an Author? In Bouchard, D. F. (Ed.), Language, Countermemory, Practice: Selected Essays and Interviews Ithaca, NY: Cornell University Press.

[10] Holton, G., & Goroff, D. (Fall 1995). "Where is American Education Going? Report on Convocation." Daedalus: Journal of the American [11] Academy of Arts and Science (American Education: Still Separate Still Unequal): pp. 1-42.

[12] Jaeger, R. M., Ed. (1998). Complementary Methods for Research in Education, Washington, D. C., American Educational Research Association.

[13] Johnson, M. (1996). Organic Education: Teaching without Failure, Montgomery, Alabama, Communication Graphics.

[14] Krippendorff, K., (1980). Content analysis: An introduction to its methodology, Beverly Hills, CA: Sage.

[15] Kruschke, E. R. &Jackson, B. M. (1987). The Public Policy Dictionary. [D]. Clio Press.

[16] Kvalseth, T. O., A Coefficient of Agreement for Nominal Scales: An Assymmetric Version of Kappa. Educational and Psychological Measurement: v51 n1 pp. 95-101 Spr. 1991.

[17] Lucas, W. F. (1983). Modeles in Applied mathematics. Vol. 2: Political and Related Models [M]. Springer Verlag, New York, Inc.

[18] Nagel, S. S. (1983), Encyclopedia of Policy Studies, Marcel Dekker, Inc. 1983.

[19] North, R. C., Holsti, M. O. Zaninovich, G., & Zinnes, D. A. (1963). Content Analysis, Evenston: Northwestern Uni. Press.

[20] Ockwell, D. (2001) A Comparative Study of Alternative Approaches to Discourse Analysis.

[21] Strobel, S. S., (1997). A Content Analysis of College and University Mission Statements (Doctoral dissertation, Temple University, 1997). Dissertation Abstracts International, 10-58, 3796.

[22] Tanner, D. & Tanner, L. (1990). History of the School Curriculum. New York: Macmillan Publishing Company.

[23] Tyler, R. W. (1949). Basic Principals of Curriculum and Instruction, Chicago, University of Chicago Press.

[24] 阿尔蒙德. 比较政治学:体系、过程、政策. 上海:上海译文出版社,1987

[25] 奥斯维特. 哈贝马斯传. 沈亚生译. 黑龙江人民出版社,1999

[26] 北京师大教育政策与法律研究所课题组. 教育研究的前沿动态和热点问题——以2002年《教育研究》杂志所发论文为例. 北京: 中国教育报, 2003-03-22

[27] 陈谭. 公共政策学. 长沙:湖南师范大学出版社,2003

[28] 村田忠禧. 从《人民日报》元旦社论看中华人民共和国的历史. 中共党史研究,2002(3)

[29] 杜刚建. 信息公开化:21 世纪的挑战和要求. 见:北京:半月谈, 2000(1-2)

[30] 樊钉. 公共政策. 北京:国家行政学院出版社,2005

[31] 房玉成,赵会民. 现代西方政治理论. 北京:北京出版社,1995

[32] 费尔克拉夫. 话语和社会变迁. 殷晓蓉译. 北京:华夏出版社,2003

[33] 福克斯,米勒. 后现代公共行政——话语指向. 楚艳红等译. 中国人民大学出版社, 2002

[34] 戈斯顿 . 公共政策的制定:程序和原理. 朱子文译. 重庆:重庆出版社,2002

[35] 桂诗春,宁春岩. 语言学方法论. 北京:外语教学与研究出版社, 1997

[36] 韩毓海. 夜阑卧听风吹雨. 北京:读书, 2001(6)

[37] 李钢 . 话语文本与国家教育政策分析(博士学位论文). 北京师范大学研究生院,2004

[38] 李岚清. 李岚清教育访谈录. 北京:人民教育出版社,2003
[39] 联合国教科文组织. 世界教育报告:2000. 北京:中国对外翻译出版公司,2001
[40] 列宁. 列宁全集(第三卷). 北京:人民出版社,1985
[41] 林德布洛姆 . 政策的制定过程. 王山译. 北京:华夏出版社,1987
[42] 刘润清. 西方语言学流派. 北京:外语教学与研究出版社, 2000
[43] 刘武俊. 压制上访违宪违法. 北京:中华工商时报, 2003-04-08
[44] 卢泰宏. 信息分析. 广州:中山大学出版社,1997
[45] 陆学艺. 当代中国社会阶层研究报告. 北京:社会科学文献出版社, 2002
[46] 罗斯金等. 政治科学(第六版). 林震等译. 北京:华夏出版社, 2001
[47] 迈尔斯. 书写生物学:科学知识的社会建构文本. 孙雍君等译. 南昌:江西教育出版社, 1999
[48] 麦克斯怀特. 公共行政的合法性——一种话语分析. 吴琼译. 中国人民大学出版社,2002
[49] 毛寿龙. 政治发展中的知识与实践的制度基础. 见:天津社会科学,2003(5)
[50] 莫谢拉. 权力的浪潮——全球信息技术的发展与前景(1964—2010). 高銛等译. 北京:社会科学文献出版社, 2002
[51] 那格尔. 政策研究百科全书. 北京:科学技术文献出版社,1990
[52] 庞景安. 科学计量研究方法论. 北京:科学技术文献出版社,1999
[53] 时殷弘. 权势的变迁——全球化和信息技术革命的权势效应. 见:学术界, 2001 (5)
[54] 史密斯. 全球化与后现代教育学. 郭洋生译,北京:教育科学出版社, 2000
[55] 斯蒂格利茨. 自由、知情权和公共话语. 北京:中国社会科学文摘, 2003(1)
[56] 斯考伦等. 跨文化交际——话语分析法. 施家炜译. 北京:社会科学文献出版社, 2000
[57] 托夫勒. 权力的转移. 刘江等译. 北京:中共中央党校出版社, 1991
[58] 杨帆. 十三届四中全会以来中国的政治发展. 见:政治学研究. 北京:中国社会科学院政治研究所,2003(3)
[59] 叶起昌. 超链接文本的话语分析. 北京师范大学博士论坛,2002-11-27
[60] 袁振国. 教育政策学. 南京:江苏教育出版社,1996
[61] 张力. 2003 年中国教育绿皮书——中国教育政策年度分析报告. 北京:教育科学出版社,2003
[62] 张力. 面对贫困:中国贫困地区教育发展的背景 · 现状 · 对策. 南宁:广西教育出版社,1998
[63] 国家教育发展研究中心. 2000 年中国教育绿皮书. 北京:教育科学出版社, 2000
[64] 赵康. 专业、专业属性及判断成熟专业的 6 条标准——一个社会学角度的分析. 北京:社会学研究, 2000(5)
[65] 中国教育与人力资源问题报告课题组. 从人口大国迈向人力资源强国. 北京:高等教育出版社, 2005